UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

DANIEL FREIRE TSUHA

Análise de vídeos de escalada de velocidade utilizando visão computacional

São Paulo
2023

DANIEL FREIRE TSUHA

Análise de vídeos de escalada de velocidade utilizando visão computacional

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Metodologia e Técnicas da Computação

Orientador: Prof. Dr. Helton Hideraldo Bíscaro

São Paulo

2023

Ficha catalográfica elaborada pela Biblioteca da Escola de Artes, Ciências e Humanidades,
com os dados inseridos pelo(a) autor(a)
Brenda Fontes Malheiros de Castro CRB 8-7012; Sandra Tokarevicz CRB 8-4936

Freire Tsuha, Daniel
Análise de vídeos de escalada de velocidade utilizando visão computacional / Daniel Freire Tsuha; orientador, Helton Hideraldo Biscaro. -- São Paulo, 2023.
101 p: il.

Dissertacao (Mestrado em Ciencias) - Programa de Pós-Graduação em Sistemas de Informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, 2023.
Versão corrigida

1. Visão Computacional. 2. Avaliação Física. 3. Escalada Esportiva. 4. Escalada de Velocidade. I. Biscaro, Helton Hideraldo, orient. II. Título.

Dissertação de autoria de Daniel Freire Tsuha, sob o título **"Análise de vídeos de escalada de velocidade utilizando visão computacional"**, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovada em 28 de fevereiro de 2023 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Helton Hideraldo Bíscaro
Universidade de São Paulo
Presidente

Prof. Dr. Roberto Hirata Junior
Universidade de São Paulo

Prof. Dr. Adriano Eduardo Lima da Silva
Universidade Tecnológica Federal do Paraná

Resumo

TSUHA, Daniel Freire. **Análise de vídeos de escalada de velocidade utilizando visão computacional**. 2023. 101 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2023.

Técnicas de visão computacional são utilizadas em diversos contextos com o objetivo de extrair informações a partir de imagens e vídeos. Uma das áreas do conhecimento que tem se beneficiado com a utilização de tais técnicas é a ciência do esporte. Essa abordagem é uma alternativa de baixo custo e não intrusiva, já que não é necessário a aquisição de nenhum equipamento adicional e nem a utilização de sensores fixados ao corpo dos atletas. A aplicação dessas técnicas em vídeos de escalada esportiva de velocidade pode ajudar profissionais da área a otimizarem os treinamentos e a detectarem pontos de melhoria dos atletas. Nesse contexto, o presente trabalho tem por objetivo validar que técnicas de visão computacional podem ser utilizadas para estimar a posição dos atletas de escalada durante provas de velocidade, de forma que os resultados obtidos sejam estatisticamente significativos se comparados com outros métodos já consolidados na área da ciência do esporte. Foi realizada uma revisão sistemática da literatura para levantar as técnicas de visão computacional mais utilizadas para a avaliação física de atletas. O próximo passo foi o desenvolvimento de um algoritmo que utiliza duas redes neurais convolucionais distintas para detectar os atletas e as agarras nas imagens. Diversas técnicas computacionais são aplicadas para combinar os dados obtidos e mapear a posição dos escaladores para dimensões do mundo real. O algoritmo foi testado em uma base de dados contendo diversas provas de escalada de velocidade e foi capaz de reconstruir a trajetória do atleta. Para validar o algoritmo um *frame* de cada vídeo da amostra ($n = 80$) foi escolhido aleatoriamente e os resultados obtidos foram comparados aos dados resultantes da medição manual. O erro médio mensurado foi de $55,5 \pm 64,9$ milímetros na medição vertical. Tal resultado demonstra a viabilidade de se utilizar técnicas de visão computacionais para se reconstruir a estimativa de atletas de escalada durante provas de velocidade. Espera-se que a ferramenta desenvolvida ajude profissionais da ciência do esporte a otimizarem o treino de atletas e que novas ferramentas de baixo custo possam ser derivadas da presente pesquisa.

Palavras-chaves: Visão Computacional, Avaliação Física, Escalada Esportiva, Escalada de Velocidade.

Abstract

TSUHA, Daniel Freire. **Analysis of speed climbing videos using computer vision**. 2023. 101 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2023.

Computer vision techniques are used in several contexts in order to extract information from images and videos. One of the areas of knowledge that has benefited from the use of such techniques is sport science. This approach is a low-cost and non-intrusive alternative, since it is not necessary to purchase any additional equipment or use sensors attached to the athletes' body. The application of these techniques in sports speed climbing videos can help professionals in the field to optimize training and detect points of improvement in athletes. In this context, the present work hypothesizes that computer vision techniques can be used to estimate the position of climbing athletes during speed climbs so that the results obtained are statistically comparable to the results of other methods already consolidated in the field of sport science. First, a systematic review of the literature was carried out to list out the computer vision techniques most used for the physical evaluation of athletes. The next step was the development of an algorithm that uses two distinct convolutional neural networks to detect athletes and holds in the images. Several computational techniques are applied to combine the obtained data and map the position of the climbers to real world dimensions. The algorithm was tested on a database containing several speed climbing events and was able to reconstruct the athlete's trajectory. To validate the algorithm, one frame of each video in the sample ($n = 80$) was randomly chosen and the results obtained were compared to data resulting from manual measurement. The mean error measured was 55.5 ± 64.9 millimeters in the vertical measurement. This result demonstrates the viability of using computational vision techniques to reconstruct the estimation of climbing athletes during sprint events. It is expected that the developed tool will help sports science professionals to optimize the training of athletes and that new low-cost tools can be derived from the present research.

Keywords: Computer Vision, Physical Assessment, Sport Climbing, Speed Climbing.

Lista de figuras

Lista de algoritmos

Lista de tabelas

Lista de abreviaturas e siglas

IFSP	International Federation of Sport Climbing (Federação internacional de escalada esportiva)
MLP	Multilayer Perceptron
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
YOLO	You Only Look Once (Nome de uma rede neural convolucional)

Lista de símbolos

Vp	Verdadeiro positivo
Fp	Falso positivo
Fn	Falso negativo
IoU	Intersection over Union (Intersecção sobre união)

Sumário

1 Introdução

Técnicas de visão computacional são utilizadas em diversos contextos com o objetivo de extrair informações a partir de imagens e vídeos. Uma das áreas do conhecimento que tem se beneficiado com a utilização de tais técnicas é a ciência do esporte. Estudos como os de Ceseracciu *et al.* (2011), Cheng, Shan e Wang (2014), Sim e Sundaraj (2010) e Shin e Ozawa (2008) apresentam soluções baseadas em visão computacional para extrair métricas de desempenho durante a prática esportiva a partir de vídeos. Essas ferramentas podem ser utilizadas por técnicos e treinadores para acompanhar a evolução do treinamento.

Entre os trabalhos desenvolvidos, encontram-se ferramentas para auxiliar no treinamento de esportes já consolidados como: natação (SHA *et al.*, 2014; VICTOR *et al.*, 2017), corrida (JINYAN; GUANLEI; YU, 2013; GADE; LARSEN; MOESLUND, 2017) e tênis (MUKAI; ASANO; HARA, 2011; SHEETS *et al.*, 2011). Com a ascensão de novos esportes, como a escalada esportiva, a demanda por novas ferramentas de apoio se torna crescente. A relevância de tal esporte é corroborada pela sua estreia como modalidade olímpica nos jogos de Tóquio 2020 (OLYMPIC, 2016). Além disso, segundo a Federação Internacional de Escalada Esportiva (IFSC), 35 milhões de pessoas escalam regularmente no mundo (DAOUST, 2018).

Na escalada esportiva os atletas utilizam agarras (pequenas peças de resina ou madeira) presas em uma parede com diversos graus de inclinação com o objetivo de alcançar o topo de uma via (rota de escalada pré-estabelecida), a figura 1 mostra as paredes e as agarras de um ginásio de escalada.

Diversas métricas de desempenho podem ser estudadas para avaliar o condicionamento dos atletas. Para extrair tais dados, várias ferramentas podem ser utilizadas para analisar o movimento dos escaladores. Em White e Olsen (2010), operadores treinados cronometravam o tempo de determinadas ações dos escaladores utilizando como referência as gravações do campeonato mundial de escalada. Em Sibella *et al.* (2007), foram utilizados marcadores reflexivos colados na pele do atleta e, durante a realização do exercício, diversas câmeras especiais capturavam o movimento dos marcadores. As imagens capturadas foram processadas por um software para recriar o movimento do escalador em três dimensões, permitindo a extração das métricas.

Figura 1 – Parede de escalada com diversos graus de inclinação e agarras.

Fonte – Daniel Freire Tsuha, 2019

Outra abordagem baseada em vídeo faz o uso de softwares como o *Kinovea* (CHARMANT, 2017). Esse tipo de ferramenta permite que um operador marque os pontos de interesse quadro a quadro durante a realização de um movimento. Feito isso, o software reconstrói a trajetória dos pontos de interesse e extrai diversas métricas cinemáticas. Esse software pode ser utilizado para analisar qualquer tipo de movimento, sendo considerado o padrão ouro para a análise de vídeo. Porém, para garantir a precisão das medições, diversos operadores devem realizar o processo de marcação, como apresentado na metodologia de Ceseracciu *et al.* (2011).

Outras técnicas podem ser utilizadas para avaliar o desempenho físico de escaladores. Em Abreu *et al.* (2019), os autores utilizaram sensores de força fixados em uma agarra para coletar os dados de desempenho. Já em Laffaye *et al.* (2014), os autores fixaram um acelerômetro à cintura dos atletas durante a realização de um exercício de escalada, com o intuito de comparar o desempenho de escaladores de modalidades distintas.

Tais ferramentas possuem limitações por demandarem equipamentos específicos e/ou mão de obra especializada para a operação. No caso da utilização de marcadores reflexivos, as câmeras possuem alto custo de aquisição e demandam mão de obra especializada para a montagem, calibração e operação do sistema. As abordagens que utilizam software como o *Kinovea* (CHARMANT, 2017) são demoradas e custosas, já que o processo é realizado de forma manual e necessita ser replicado por diversos operadores treinados. Já a utilização de

sensores demanda a aquisição de aparelhos específicos, que necessitam de pessoas treinadas para operar os equipamento e processar os dados coletados.

Na modalidade de escalada de velocidade, em que o objetivo dos atletas é escalar uma parede padronizada no menor tempo possível, estimar a posição dos escaladores durante a prova permite que treinadores analisem o desempenho do atleta em cada seção da corrida. Durante o processo de revisão bibliográfica, não foram encontrados estudos que utilizem técnicas de visão computacional para avaliar de desempenho físico de escaladores. Entretanto, em uma publicação recente em Pandurevic *et al.* (2022) apresentou=se uma ferramenta para analisar o desempenho de escaladores de velocidade. O presente trabalho possuí intersecções com a metodologia apresentada em Pandurevic *et al.* (2022), entretanto a principal diferença está na metodologia utilizada para estimar a posição dos escaladores e na avaliação da precisão da ferramenta.

Ainda que já existam estudos relacionados à escalada esportiva, a revisão sistemática (capítulo 3) demonstrou ser um campo com diversas lacunas à serem preenchidas e a utilização de técnicas de visão computacional podem auxiliar no processo de avaliação física a partir de vídeos. Com a popularização da transmissão de eventos esportivos, tais técnicas possibilitam a extração automática dos dados, sem a necessidade de operadores previamente treinados para utilizar o sistema. Além disso, também dispensa a aquisição de equipamentos específicos de avaliação física, podendo ser aplicados à competições já realizadas.

1.1 Hipótese e objetivo

Após identificar as lacunas e oportunidades na literatura, o presente trabalho tem por hipótese de que técnicas de visão computacional podem ser utilizadas para estimar a posição dos atletas de escalada durante provas de velocidade de forma que os resultados obtidos sejam estatisticamente significativos se comparados com outros métodos já consolidados na área da ciência do esporte.

Para testar tal hipótese, o objetivo dessa pesquisa é desenvolver e validar uma ferramenta baseada em visão computacional capaz de estimar a posição do centro de massa de atletas de escalada de velocidade a partir de gravações de campeonatos oficiais disponíveis na internet. Espera-se que o método desenvolvido seja tão preciso quanto outros métodos já utilizados na área da ciência do esporte.

Os objetivos específicos do trabalho são:

- implementar um sistema capaz de estimar a posição dos atletas durante provas de escalada esportiva;
- validar estatisticamente a precisão do sistema comparando os resultados com técnicas já consolidadas na área de educação física, como a marcação manual dos pontos de interesse.

1.2 Justificativa

Diversas ferramentas baseadas em visão computacional foram desenvolvidas para auxiliar no treinamento esportivo, entretanto, a quantidade de pesquisas que utilizam tais algoritmos para auxiliar no treinamento de praticantes de escalada esportiva ainda é baixo se comparado com outros esportes.

Sendo assim, caso a hipótese se demonstre verdadeira, escaladores poderão se beneficiar com uma ferramenta de fácil utilização e que não utilize equipamentos de medições específicos. Além disso, espera-se que a ferramenta desenvolvida sirva como base para a implementação de outros sistemas de avaliação física.

O intuito da presente pesquisa é preencher uma lacuna da ciência do esporte e contribuir para a área de visão computacional utilizando dados já disponíveis, porém pouco utilizados. Além disso espera-se contribuir para a automatização de uma tarefa que outrora seria realizada de forma manual.

1.3 Resumo da metodologia

As etapas que foram desenvolvidas para alcançar o objetivo e testar a hipótese são:

- **Criação de um conjunto de dados:** foi criado um conjunto de vídeos contendo 80 corridas de quatro competições oficiais de escalada de velocidade. Além disso, foi criado um conjunto de dados contendo 7.165 agarras rotuladas distribuídas em mil *frames*;
- **Mapeamento das agarras:** o passo seguinte foi treinar uma rede neural convolucional para detectar as agarras, para essa tarefa foi utilizada a rede *YOLO*;

- **Identificação das agarras:** após detectar a posição das agarra, é necessário identifica-las para estimar a posição dos atletas na parede;
- **Detecção dos escaladores:** para obter a posição dos escaladores nas imagens foi utilizada a rede neural convolucional *OpenPose*, que foi desenvolvida especificamente para detectar e esqueletizar humanos em vídeos;
- **Estimativa de posição dos escaladores:** nesta etapa é realizada a conversão das medidas da imagem para as medidas do mundo real;
- **Validação:** para validar a confiabilidade da ferramenta desenvolvida, comparou-se os resultados obtidos de forma automática com as marcações realizadas de forma manual.

A metodologia completa está descrita no capítulo 4.

1.4 Organização do trabalho

O restante desse documento está estruturado da seguinte forma: no capítulo 2 são apresentados os conceitos fundamentais necessários para o entendimento do trabalho; o capítulo 3 apresenta a revisão sistemática que contempla estudos semelhantes, porém com aplicação em outros esportes; o capítulo 4 detalha a metodologia utilizada para o desenvolvimento da pesquisa; o capítulo 5 apresenta os resultados obtidos; e por fim, o capítulo 6 apresenta as conclusões da pesquisa.

2 Conceitos fundamentais

O presente capítulo tem por objetivo apresentar os conceitos utilizados no desenvolvimento da pesquisa. Sendo a escalada esportiva de velocidade o objeto de análise, a seção 2.1 tem por objetivo descrever a parede utilizada durante as provas. As seções seguintes descrevem os conceitos computacionais aplicados para a realização da análise. A seção 2.2 descreve formalmente imagens digitais e como são representadas em sistemas computacionais. A seção 2.3 apresenta o conceito de aprendizagem de máquina, um conjunto de técnicas para realizar predições a partir da análise de um conjunto de dados. Já a seção 2.4 apresenta as redes neurais artificiais, um conjunto de algoritmos que mimetiza o funcionamento de neurônios. Por fim, a seção 2.5 descreve as redes neurais convolucionais, um tipo de rede especialmente útil para o processamento de imagens.

2.1 Escalada de velocidade

A escalada de velocidade é uma das três disciplinas da escalada esportiva em que o objetivo é alcançar o topo da via o mais rápido o possível. A parede de escalada dessa modalidade é padronizada pela Federação Internacional de Escalada Esportiva (IFSC) e as agarras, peças utilizadas pelos atletas para escalar, também possuem formato e posições definidas para esse tipo de competição. O restante desta seção se baseia nos documentos oficiais disponíveis no *site* oficial da instituição[1,2].

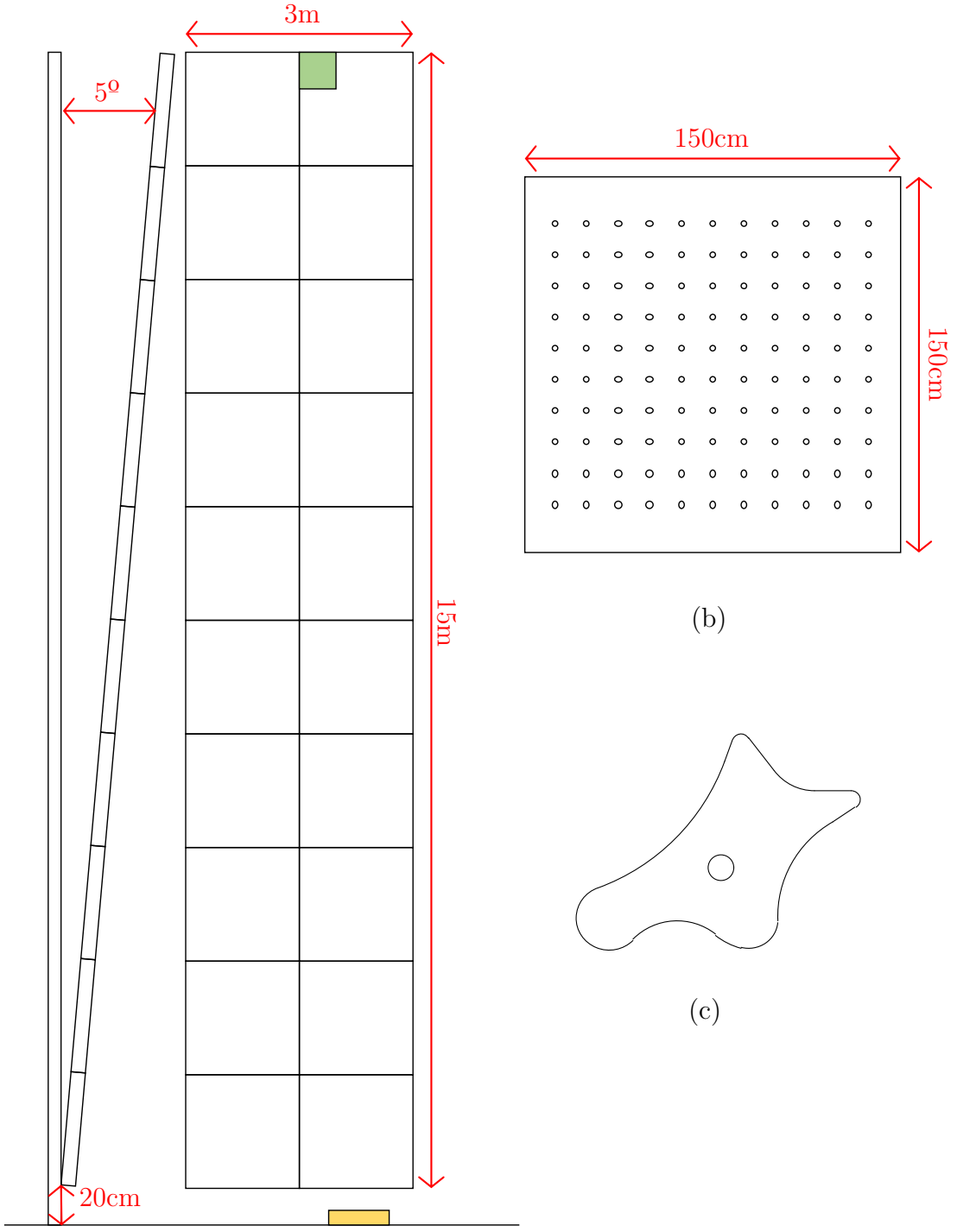
A parede de escalada de velocidade possui quinze metros de altura por três de largura e é formada por vinte painéis de 150 x 150 cm onde as agarras são fixadas. Os painéis devem possuir furos equidistantes a cada 125 milímetros, as marges superiores e inferiores devem ser de 187,5 mm enquanto as margens laterais são de 125 mm. A inclinação das placas deve ser de 5° em relação à parede de sustentação. Além disso também são utilizados sensores para detectar a partida e a chegada dos atletas. A figura 2 mostra a estrutura da parede.

A posição e inclinação das agarras também são padronizadas, a ponta oposta à parte arredondada deve estar alinhada com um dos furos da parede como descrito na

[1] Especificações da parede disponível em ⟨https://cdn.ifsc-climbing.org/images/ifsc/Footer/Manufacturers/Speed_Licence_Rules_Walls.pdf⟩. Acessado em 09/01/2023.

[2] Regras da competição disponível em ⟨https://www.ifsc-climbing.org/index.php/world-competition/rules⟩. Acessado em 09/01/2023.

Figura 2 – Especificação da parede de escalada: (a) Medidas da parede de escalada, o retângulo amarelo indica a posição do sensor de partida e o quadrado verde a posição do sensor de chegada; (b) painéis utilizados para fixar as agarras; (c) formato da agarra utilizada nas competições.



(a)

(b)

(c)

Fonte – Daniel Freire Tsuha, 2023

documentação oficial. A figura 3 mostra a posição das agarras definida pelo IFSC. Para facilitar o entendimento, no presente trabalho as agarras foram nomeadas de baixo para cima utilizando as letras do alfabeto. No restante desse trabalho, tal notação será utilizada para identificar cada uma das agarras.

2.2 Imagem digital

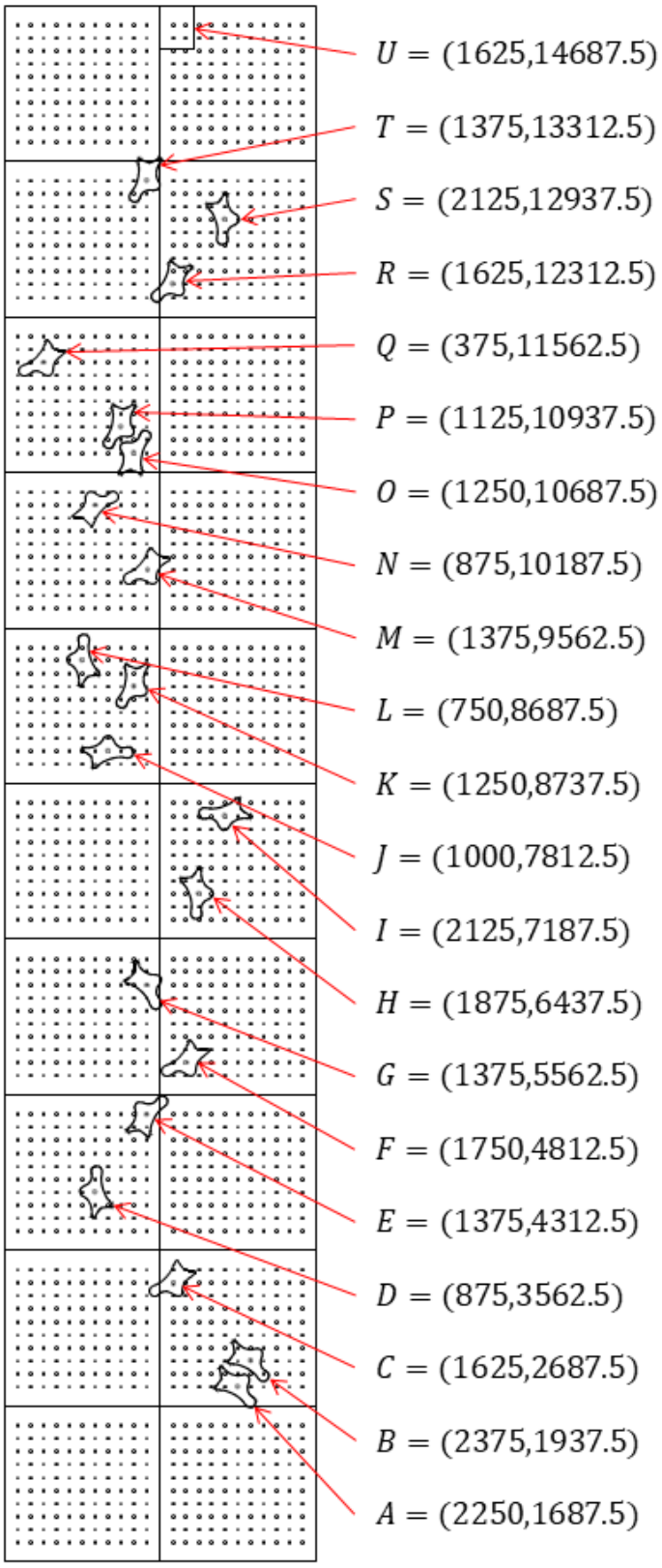
De maneira formal, "uma imagem pode ser definida como uma função bidimensional $f(x, y)$, em que x e y são coordenadas espaciais" (GONZALEZ; WOODS, 2011) e f representa a intensidade de um sinal para quaisquer valores de (x, y) (GONZALEZ; WOODS, 2011). Esses sinais podem representar a refletância de um objeto ao interagir com uma fonte luminosa, os níveis de raios-x que atravessam os tecidos do corpo humano ou sinais captados por equipamentos de ressonância magnética (GONZALEZ; WOODS, 2011).

Ballard e Brown (1982) definem uma imagem como um modelo geométrico bidimensional que descreve a projeção de uma cena tridimensional, e sua formação se dá pelo registro da radiação que incide sobre objetos físicos. A refletância capturada a partir da interação de uma fonte luminosa com os objetos da cena pode ser entendida como o "brilho" de um objeto dado sua geometria e características intrínsecas (BALLARD; BROWN, 1982; GONZALEZ; WOODS, 2011).

Uma imagem é considerada digital se f, x e y assumirem valores finitos e discretos (BALLARD; BROWN, 1982; GONZALEZ; WOODS, 2011). Uma das formas de aquisição desse tipo de imagem são as câmeras digitais. Nesses equipamentos uma lente ótica é utilizada para projetar a cena em um sensor de modo que este coincida com o plano focal da lente (GONZALEZ; WOODS, 2011).

Os sensores de captura utilizados em câmeras digitais são formados por unidades sensoras que normalmente são dispostas em formato retangular (GONZALEZ; WOODS, 2011) e a imagem capturada é uma amostra regularmente espaçada da cena (BALLARD; BROWN, 1982). Além disso, o formato do sensor define a amostragem espacial (BURGER; BURGE, 2016) e, consequentemente, o formato da imagem. Cada unidade sensora é capaz de transformar energia luminosa em tensão elétrica proporcional à quantidade de luz captada. Os sinais elétricos são então transformados em sinais digitais que possuem

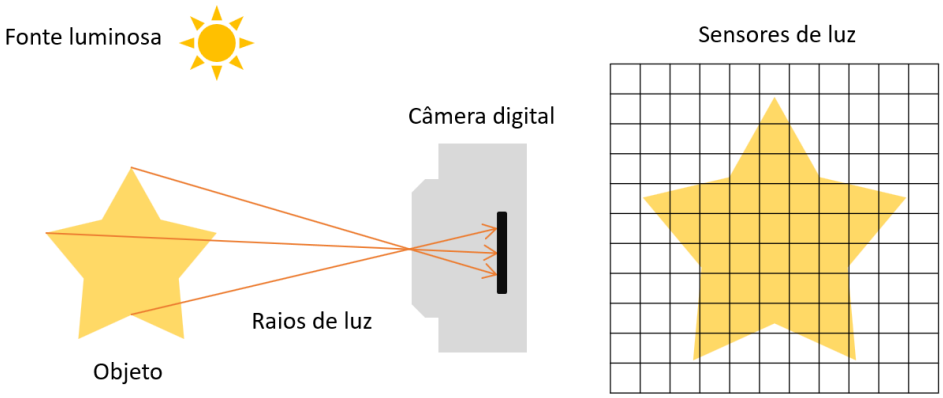
Figura 3 – Posição das agarras da parede de escalada em milímetros: foi considerado a posição do parafuso de fixação central como referência.



Fonte – Daniel Freire Tsuha, 2023

um intervalo finito de possíveis valores, esse processo é conhecido como discretização (GONZALEZ; WOODS, 2011; BURGER; BURGE, 2016). A figura 4 ilustra o processo de captura e geração de uma imagem digital.

Figura 4 – Exemplo de sensor utilizado em um câmera digital: os raios de luz emitido por uma fonte luminosa são refletidos pelo objeto. Os raios que chegam à lente da câmera são direcionados para um componente eletrônico contendo diversos sensores de luz.



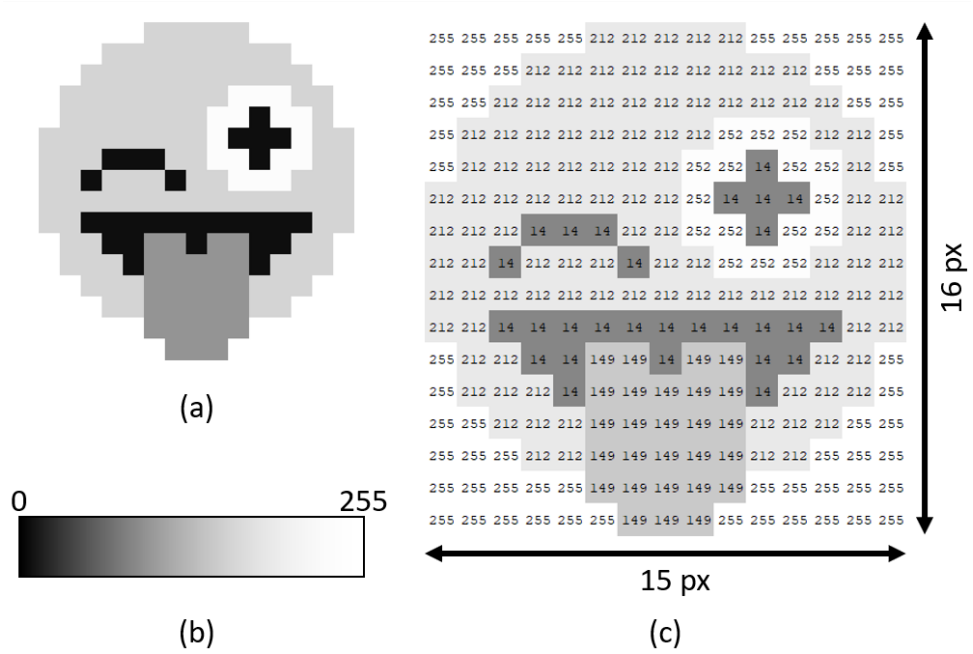
Fonte – Daniel Freire Tsuha, 2023

Computacionalmente, uma imagem é formada por elementos pictóricos de tamanho finito (conhecidos como *pixels*) armazenados em uma matriz bidimensional. O tamanho de uma imagem é definido pela quantidade de linhas e colunas da matriz (quantidade de *pixels*), já a resolução espacial define a medida do menor nível de detalhe de uma imagem no mundo real. Para impressoras, é comum se utilizar a medida de dpi (*dots per inch* ou pontos por polegada), já imagens de satélite podem utilizar medidas como *pixels* por quilômetro para definir a resolução de fotos aéreas (BURGER; BURGE, 2016; BALLARD; BROWN, 1982; GONZALEZ; WOODS, 2011).

A resolução de contraste é o valor que os *pixels* podem assumir e representam a "menor variação discernível de nível de intensidade na imagem" (GONZALEZ; WOODS, 2011). Normalmente esses valores são representados por inteiros no intervalo de 0 até $2^k - 1$, sendo k um valor inteiro (GONZALEZ; WOODS, 2011). No caso de imagens em tons de cinza, uma única matriz é utilizada para representar a intensidade do brilho dos *pixels*. A figura 5 resume os conceitos apresentados até o momento.

Para representar imagens coloridas, são utilizados espaços de cores, que são "uma maneira intuitiva de organizar as cores percebidas por humanos" (BALLARD; BROWN, 1982). Na prática, o modelo de cor mais utilizado por hardwares como monitores e câmeras é o RGB (*Red, Green, Blue*) (GONZALEZ; WOODS, 2011; BURGER; BURGE, 2016).

Figura 5 – Exemplo de imagem digital em tons de cinza: (a) Imagem original em tons de cinza; (b) a resolução de contraste é inteiro entre zero e 255 que indica a quantidade de tons disponíveis; (c) uma matriz bidimensional em que cada célula indica a tonalidade de cinza de um determinado *pixel* (px), a quantidade de linhas e colunas indicam a resolução espacial da imagem.



Fonte – Daniel Freire Tsuha, 2023

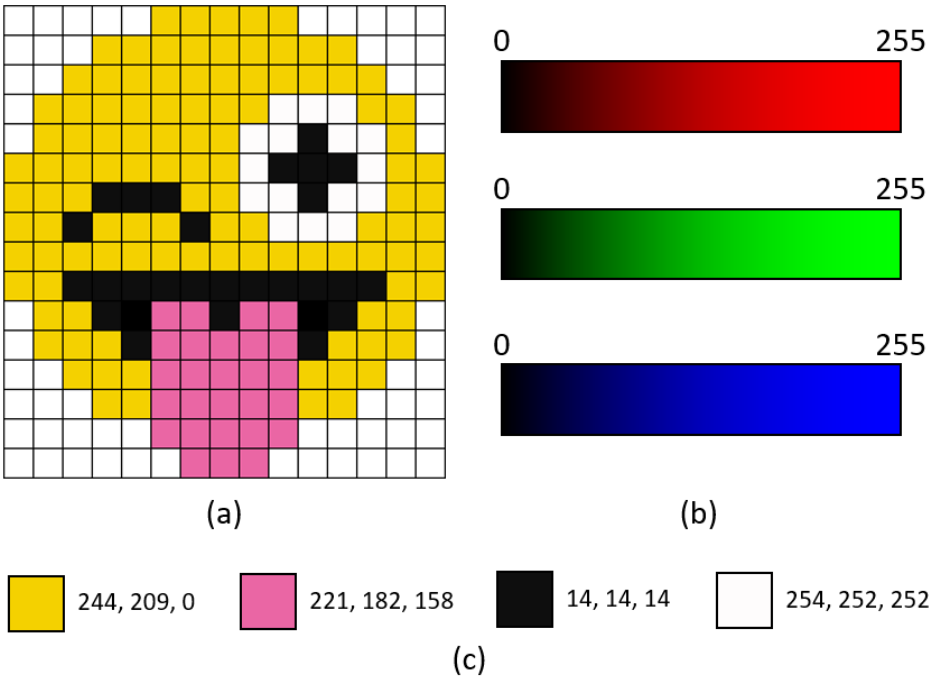
Para representar os níveis de intensidade de cada cor, utiliza-se três matrizes de intensidade, também chamadas de canais. (BURGER; BURGE, 2016). A figura 6 mostra uma imagem colorida que utiliza o modelo de cor RGB.

2.3 Aprendizado de máquina

De acordo com Mohri, Rostamizadeh e Talwalkar (2012), aprendizado de máquina pode ser definido como um conjunto de algoritmos capazes de utilizar dados previamente coletados para realizar predições. Tais algoritmos combinam técnicas de computação, estatística, probabilidade e otimização para criar modelos que permitam realizar tarefas como classificação e agrupamento para novos dados.

Para Bonaccorso (2017) o aprendizado de máquina consiste no desenvolvimento de modelos matemáticos capazes de realizar inferências sem o conhecimento de todos os elementos disponíveis. São algoritmos capazes de generalizar regras e aprender suas estruturas com precisão relativamente alta.

Figura 6 – Exemplo de imagem digital colorida: (a) Imagem original colorida; (b) a cor de cada *pixel* é definida pela combinação de diferentes valores de vermelho, verde e azul; (c) exemplos de cores geradas a partir da combinação dos canais RGB.



Fonte – Daniel Freire Tsuha, 2023

Diferente de outras classes de algoritmos que são avaliados em termos de complexidade de tempo e espaço, algoritmos de aprendizado de máquina também são avaliados em termos de complexidade de amostras. A complexidade de amostras diz respeito à quantidade de dados (exemplos) necessários para que o algoritmo aprenda uma família de conceitos. Sendo assim, a qualidade e a quantidade dos dados utilizados como entrada influenciam diretamente na capacidade preditiva do modelo (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

Ainda segundo Mohri, Rostamizadeh e Talwalkar (2012), as principais tarefas realizadas por esses algoritmos são:

- **Classificação:** capacidade de predizer a qual classe um item pertence (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012) (e.g. classificar a qual categoria pertence uma notícia (política, economia, esportes, etc.));
- **Regressão:** capacidade de atribuir um valor real para cada item, também entendido como o estudo da relação de dependência entre duas variáveis (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012) (e.g. quantidade de acessos à um determinado site durante um dia e horário específico);

- **Ranqueamento:** capacidade de ordenar uma lista de item de acordo com algum critério (e.g. ordenação dos resultados de um buscador de sites) (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012);
- **Agrupamento:** capacidade de particionar um conjunto de itens de forma a criar subgrupos homogêneos (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012) (e.g. identificar grupos de clientes com comportamentos similares a partir de suas compras);

2.3.1 Treinamento

Dependendo do tipo de tarefa a ser realizada, os algoritmos de aprendizagem de máquina utilizam diferentes estratégias de treinamento. Dentre as principais técnicas, pode-se citar o aprendizado supervisionado, o aprendizado não supervisionado e o aprendizado por reforço (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012; BONACCORSO, 2017; KUBAT, 2017). O aprendizado supervisionado, que será utilizado nesse trabalho para detectar as agarras e os escaladores, consiste em apresentar ao algoritmo um conjunto de dados previamente rotulados e, a partir de cada amostra, uma dada função de erro deve ser minimizada. Em outras palavras, para cada dado apresentado, os parâmetros internos do algoritmo são ajustados de forma a aprender como classificar os dados (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012; BONACCORSO, 2017).

Já no aprendizado não supervisionado, os dados de entrada não necessitam estar previamente rotulados. Nessa abordagem não há um supervisor ou medida absoluta de erro. O foco desses algoritmos é extrair propriedades úteis de como os dados se organizam para entender o comportamento deles. A principal tarefa de aprendizagem não supervisionada é o agrupamento (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012; BONACCORSO, 2017; KUBAT, 2017).

No aprendizado por reforço, o processo de treinamento e teste acontecem de forma intercalada. Nessa estratégia, o algoritmo interage com o ambiente e, de acordo com suas ações recebe uma recompensa ou penalidade, que é utilizada para ajustar os parâmetros internos do algoritmo de forma a maximizar as recompensas (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012; BONACCORSO, 2017; KUBAT, 2017). Bonaccorso (2017) utiliza como exemplo de aprendizado por reforço um algoritmo capaz de aprender a interagir com jogos eletrônicos a partir de critérios como maximizar os pontos obtidos e evitar derrotas.

Em relação ao processo de aprendizagem de um algoritmo supervisionado, Mohri, Rostamizadeh e Talwalkar (2012) resume o processo nas seguintes etapas:

- **Particionamento dos dados:** Uma das características importantes de um algoritmo de aprendizagem de máquina é a capacidade de rotular dados não vistos anteriormente (i.e. generalização). Para garantir que o algoritmo não está viciado na amostra de treinamento (sobreajuste), o conjunto de dados é dividido em duas partições: treinamento e teste. Os dados de treinamento são utilizados para "ensinar" o algoritmo durante a fase de treinamento. Já os dados de teste são utilizados para testar a capacidade modelo de realizar predições para dados não visto previamente e calcular métricas de erro (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012; BONACCORSO, 2017; KUBAT, 2017);
- **Seleção de características:** Em um conjunto de dados rotulados, nem todas as informações presentes são relevantes para o treinamento do algoritmo. A utilização de dados irrelevantes pode afetar drasticamente os resultados. Além disso, o uso de parâmetros que possuem alta correlação não melhoram a capacidade preditiva e tornam o modelo mais complexo em termos de número de parâmetros. Nesse contexto, o conhecimento a priori de um especialista é necessário para selecionar as informações mais relevantes (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012; BONACCORSO, 2017; KUBAT, 2017);
- **Otimização de hiperparâmetros:** Também conhecidos como parâmetros livres, os hiperparâmetros são responsáveis por controlar o processo de aprendizagem. Tais parâmetros influenciam diretamente no desempenho do modelo e, para otimizá-los, e necessário treinar o modelo com diferentes configurações. Nessa etapa utiliza-se um subconjunto dos dados para reduzir o custo computacional dos testes (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012; MICROSOFT, 2022);
- **Treinamento:** Na etapa de treinamento, os dados de treinamento são apresentados ao algoritmo. Para cada valor, o algoritmo testa uma hipótese e compara a saída obtida com a saída esperada. A partir dessas comparações, o modelo ajusta os seus parâmetros livres internos de forma a minimizar o valor global da função de erro e, consequentemente, aumentar a precisão dos resultados (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012; BONACCORSO, 2017);

- **Avaliação:** Por último, os dados de teste são apresentados ao algoritmo para mensurar a capacidade preditiva do modelo. Considerando a quantidade de erros e acertos, pode-se calcular métricas como precisão e acurácia (KUBAT, 2017).

2.4 Redes neurais artificiais

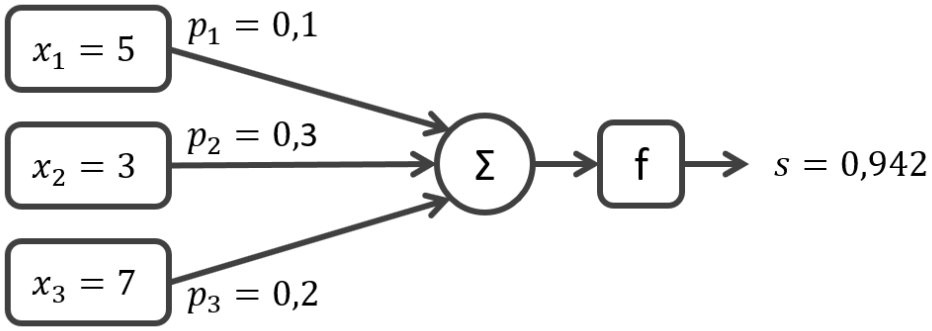
Redes neurais artificiais são uma classe de algoritmos de aprendizagem de máquina que possuem como objetivo simular o processo de aprendizagem de seres vivos. Tais algoritmos utilizam diversas unidades de processamento, conhecidas como neurônios artificias, que mimetizam a propagação dos sinais entre neurônios biológicos (ANDERSON, 1995; AGGARWAL, 2018). O primeiro modelo de neurônio artificial foi proposto em McCulloch e Pitts (1943).

De forma simplificada, um neurônio biológico possuí dendritos, corpo celular e axônios. Os dendritos são responsáveis por receber os sinais de outros neurônios e conduzi-los ao corpo celular. Se a soma de todos os sinais recebidos ultrapassar um determinado limiar, então o corpo celular irá disparar um sinal e o axônio irá propagá-lo para outros neurônios. A intensidade dos sinais propagados podem ser alterados por estímulos externos e essa mudança leva ao aprendizado em seres vivos (ANDERSON, 1995; AGGARWAL, 2018; KELLEHER, 2019).

Em um neurônio artificial, também conhecido como *perceptron*, cada dado de entrada é multiplicado por um peso sináptico e a soma desses valores é utilizado como parâmetro de entrada em uma função de ativação, assim como em um neurônio biológico. Tais funções são usadas para definir a intensidade do sinal propagado. Diversas funções podem ser utilizadas para ativação dos neurônios dependendo do tipo de problema a ser resolvido (AGGARWAL, 2018; GURNEY, 2018; SKANSI, 2018). A figura 7 sintetiza o funcionamento de um neurônio artificial e a figura 8 mostra alguns exemplos de funções de ativação.

Uma rede neural artificial é composta por diversos neurônios artificiais organizados em camadas e conectados entre si. Dependendo do tipo de problema a ser resolvido, pode-se organizar os neurônios e as conexões de diversos modos. A seguir, será considerada a arquitetura de uma rede *Multilayer Perceptron* (MLP) para descrever os conceitos básicos de uma rede neural artificial.

Figura 7 – Neurônio artificial: cada valor x_n representa uma entrada que deve ser multiplicada pelo respectivo peso p_n, os resultados são somados e utilizados como parâmetro de entrada da função f (1). Neste exemplo foi utilizada a função logistíca (2).

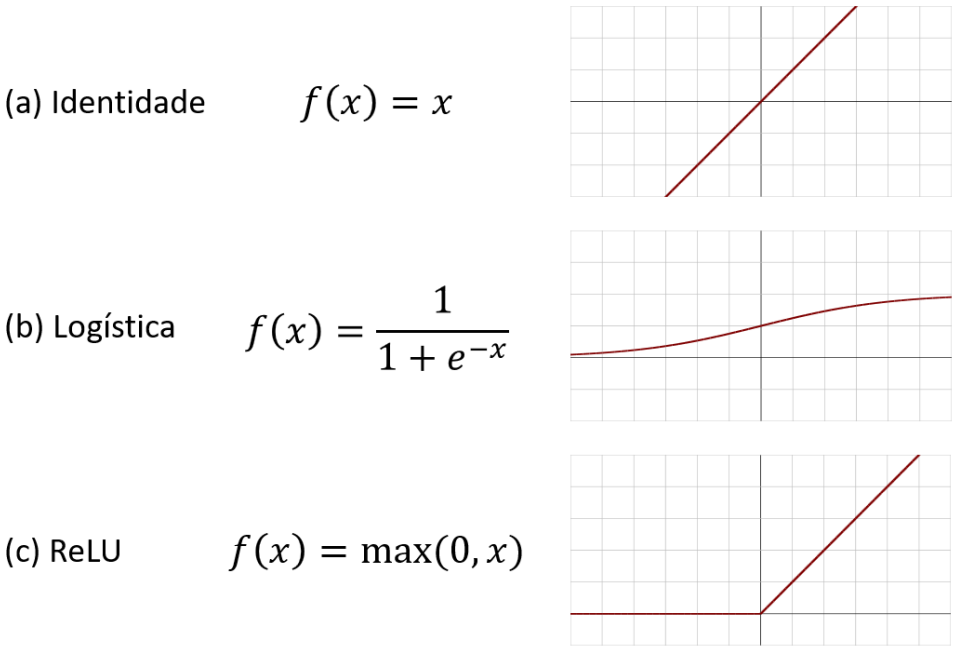


$$1)\ (5 * 0{,}1) + (3 * 0{,}3) + (7 * 0{,}2) = 2{,}8$$

$$2)\ \frac{1}{1 + e^{-2{,}8}} = 0{,}942$$

Fonte – Daniel Freire Tsuha, 2023

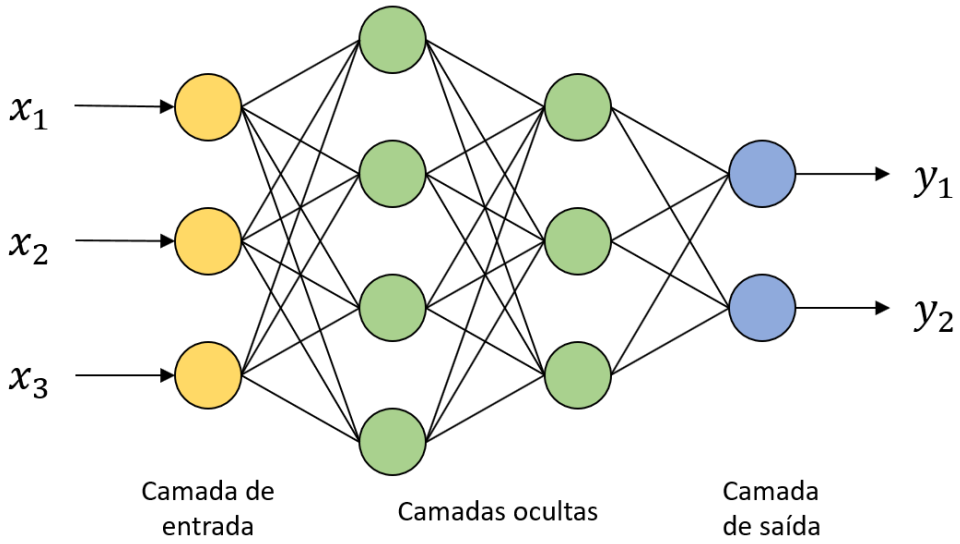
Figura 8 – Exemplos de funções de ativação utilizadas em neurônios artificiais.



Fonte – Adaptado de Laughsinthestocks em Wikimedia Commons

Em uma rede neural artificial, a primeira camada tem a função de apenas receber os dados de entrada e propagá-los para a próxima camada da rede. A quantidade de neurônios nessa camada deve ser igual a quantidade de características que descrevem o objeto a ser classificado (SKANSI, 2018). Já a quantidade de neurônios da última camada depende do tipo de problema a ser resolvido. Se o objetivo da rede é realizar uma regressão, então a última camada possuirá apenas um neurônio de saída; já no caso em que o objetivo da rede é classificar os dados, então a quantidade de neurônios na última camada será igual à quantidade de possíveis classes na maioria dos casos (AGGARWAL, 2018). As camadas intermediárias, também chamadas de camadas ocultas, podem possuir qualquer número de neurônios. A definição da quantidade de camadas e neurônios nas camadas ocultas devem ser ajustados durante o processo de otimização de hiperparâmetros (KELLEHER, 2019). A figura 9 mostra um exemplo de rede neural.

Figura 9 – Exemplo de uma rede neural com três neurônios na camada de entrada, duas camadas ocultas contendo quatro e três neurônios respectivamente e dois neurônios de saída (cada neurônio de saída representa uma classe). x_i representa as características do objeto a ser classificado e cada valor de y_j representa o pertencimento do objeto a cada classe (quanto maior o valor, mais provável que o objeto pertença àquele grupo).

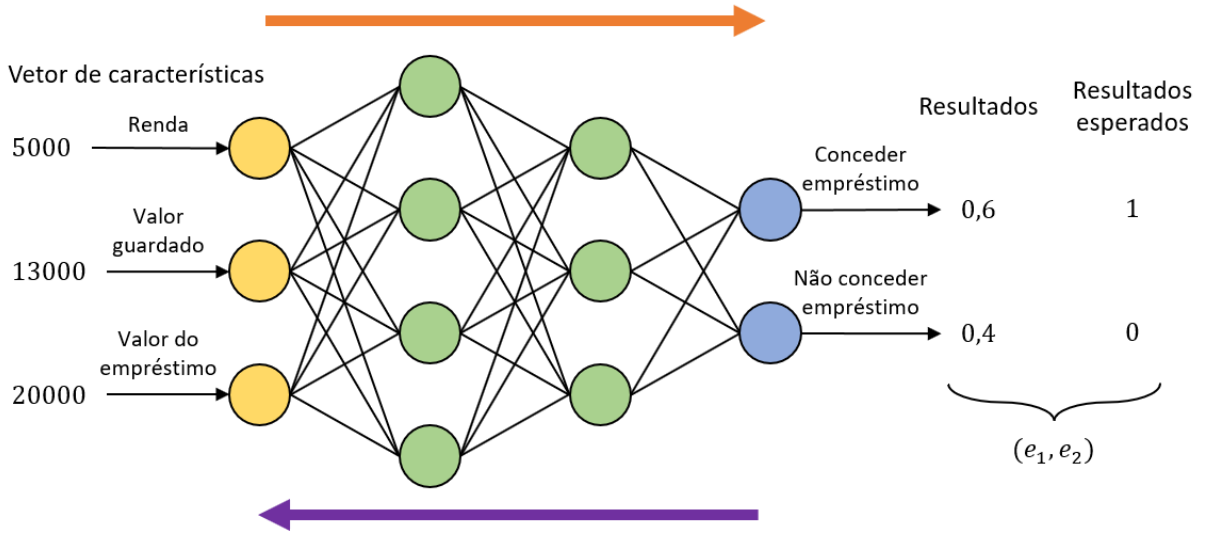


Fonte – Daniel Freire Tsuha, 2023

De modo geral, os pesos sinápticos de uma rede neural são aprendidos durante a fase de treinamento utilizando o algoritmo de *backpropagation*, ou retropropagação (AGGARWAL, 2018). Nessa etapa, o conjunto de treinamento é apresentado à rede e, a partir do erro computado entre a saída obtida e a saída espera, os pesos são ajustados de forma iterativa. O erro da rede é dado por uma função de perda que ao ser minimizada,

faz com que a rede aprenda a classificar os dados corretamente. Para definir o novo valor do peso sináptico, o algoritmo calcula o gradiente descente da função de perda e ajusta os valores da última para a primeira camada (KUBAT, 2017; AGGARWAL, 2018). A figura 10 ilustra o processo de treinamento de uma RNA.

Figura 10 – Exemplo de treinamento de uma RNA utilizando o algoritmo *backpropagation*: neste exemplo uma rede recebe como parâmetros de entrada um vetor de características contendo informações de um cliente que solicita um empréstimo e a tarefa da rede é aprovar ou não a operação. A seta laranja indica a fase de propagação em que os dados são processados pela rede. Ao final dessa etapa, o resultado é comparado com a saída esperada (no caso, o empréstimo deve ser aprovado). O erro da rede é calculado e um vetor contendo os erros de cada uma das saídas é gerado, então esse vetor é utilizado na fase de retropropagação para atualizar os pesos sinápticos (seta roxa). O processo se repete para todos os dados diversas vezes.



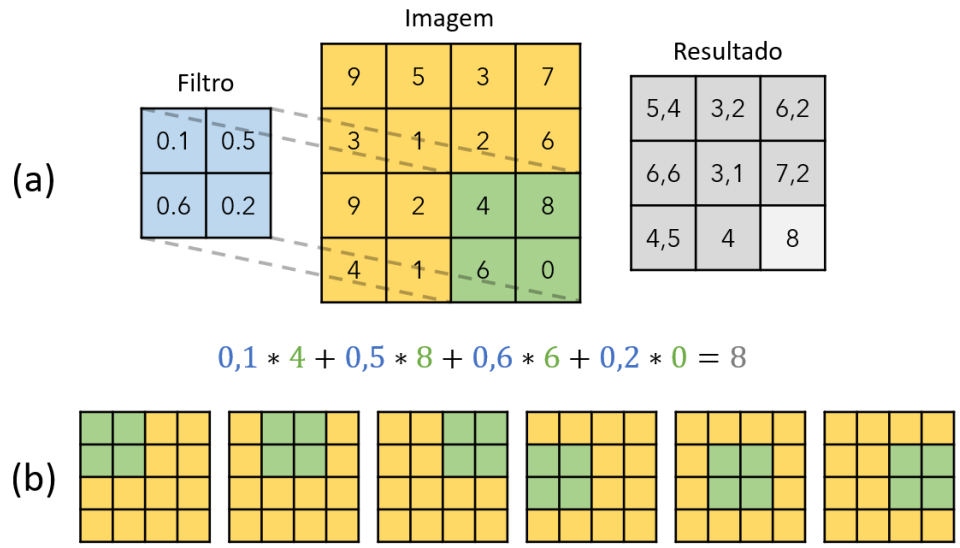
Fonte – Daniel Freire Tsuha, 2023

2.5 Redes neurais convolucionais

Redes neurais convolucionais (RNC) foram apresentadas em LeCun *et al.* (1989) com o objetivo de identificar os dígitos de códigos postais escritos à mão. Tais redes foram desenvolvidas para trabalhar em estrutura de dados no formato de grade, de forma que as primeiras camadas sejam capazes de extrair características visuais da imagem para que os neurônios das última camada sejam capazes de combiná-los e classificá-los (AGGARWAL, 2018; KELLEHER, 2019).

Para que uma RNC consiga extrair tais características visuais de uma imagem, as primeiras camadas realizam operações de convolução. Tal operação consiste em deslizar um filtro sobre a imagem realizando a operação de produto escalar de forma que o resultado seja um mapa contendo as características relevantes. Os filtros são aprendidos durante a fase de treinamento, e a quantidade dos mesmos é um hiperparâmetro da rede (AGGARWAL, 2018; KELLEHER, 2019). A figura 11 mostra um exemplo desse processo.

Figura 11 – Convolução: a) um exemplo numérico do produto escalar do filtro com a imagem; b) exemplos de posições do filtro deslizando sobre a imagem.

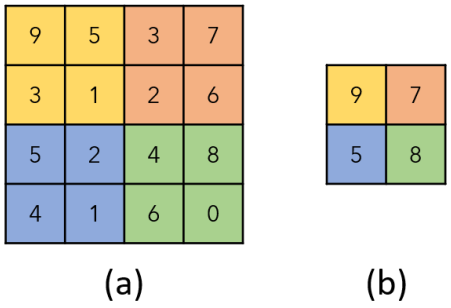


Fonte – Daniel Freire Tsuha, 2023

Após a computação dos mapas de característica, o próximo passo é a aplicação da função de ativação. No caso de RNCs, a função $ReLU$ é frequentemente utilizada para cada posição do mapa, já que sua utilização aumenta significativamente a precisão das redes. O passo seguinte no processamento de uma RNC é a operação de agrupamento, também conhecida como *pooling*. Nessa operação, uma região do mapa de característica é comprimida em um único ponto, gerando assim um novo mapa de característica. Essa operação é utilizada para reduzir a quantidade de dados. A função mais utilizada é a *max-pooling*, que tem por objetivo pegar o maior valor de uma determinada região (AGGARWAL, 2018; SKANSI, 2018; KELLEHER, 2019). A figura 12 mostra um exemplo desse processo.

Ao final do processamento dos mapas de características, tais dados são processados por uma rede completamente conectada, que utiliza a arquitetura de uma MLP nas arquiteturas tradicionais. Nessa camada, os neurônios aprendem a como integrar as

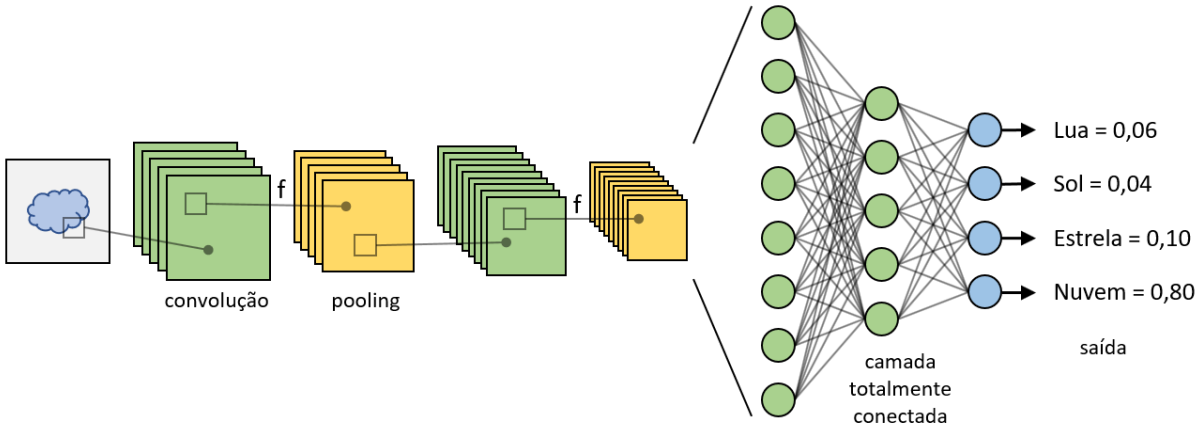
Figura 12 – Exemplo de aplicação da operação de *max-pooling* em uma imagem: a) mapa original; b) resultado da operação.



Fonte – Daniel Freire Tsuha, 2023

informações de diferentes filtros e realizar a classificação. Dependendo da arquitetura da rede, tais processos podem ser combinados de diferentes formas para se obter melhores resultados (AGGARWAL, 2018; KELLEHER, 2019). A figura 13 mostra um exemplo de uma rede neural convolucional.

Figura 13 – Rede neural convolucional: na primeira parte de uma rede neural convolucional, são aplicadas as operações de convolução juntamente com função de ativação e *pooling*, a segunda parte consiste em uma rede totalmente conectada que retorna a probabilidade de uma imagem pertencer à uma classe.



Fonte – Daniel Freire Tsuha, 2023

Redes neurais convolucionais também pode ser utilizadas para solucionar problemas como análise de texto e séries temporais (AGGARWAL, 2018). Porém, no contexto do presente trabalho, tais redes foram utilizadas para detectar as agarras e os atletas durante as competições de escalada de velocidade. A subseção 2.5.1 apresenta a rede *YOLO*, utilizada para detectar as agarras e a subseção 2.5.2 apresenta a rede *OpenPose*, utilizada para detectar os escaladores.

2.5.1 You Only Look Once

A rede neural convolucional *YOLO* (*You Only Look Once*) foi apresentada em Redmon *et al.* (2016), e diferentemente de outras RNC utilizadas para detectar objetos, tal rede divide a imagem em um padrão quadriculado e analisa cada sub-região de forma independente. Tal otimização permite que a *YOLO* processe até 45 quadros por segundo[3]. A abordagem utilizada permite que a rede processe toda a imagem de uma única vez e generalize a representação de objetos, diferente de outras RNCs que analisam porções da imagem individualmente.

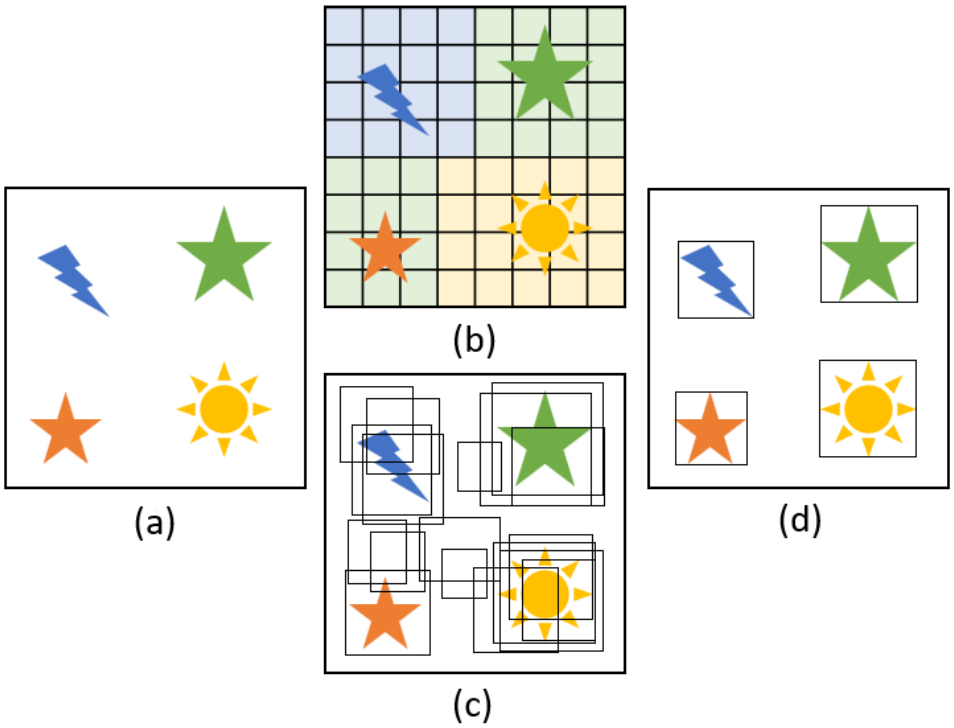
O funcionamento da rede se dá em duas partes: localização de objetos e classificação. Na primeira parte, o algoritmo procura por possíveis regiões que contenham objetos independentemente de suas classes. Para cada região candidata, a rede retorna cinco dados: as coordenadas do centro do objeto (x, y), a altura e largura (h, w), a confiança de que há um objeto naquela região (c). Tais regiões são chamadas de *bounding box*; Já na segunda parte, cada região quadriculada recebe um rótulo que descreve o conteúdo da região. O rótulo indica que se há um objeto nessa região, tal objeto pertence a determinada classe (REDMON *et al.*, 2016).

Após o cálculo de ambas as etapas, as informações são combinadas para detectar os objetos na imagem. Cada região quadriculada é responsável por classificar um determinado número de *bouding boxes* utilizando como referência o centro da região candidata. Se a confiança de um *bouding box* for maior que um determinado limiar, então o algoritmo retorna as coordenadas da região juntamente com a classe do mesmo (REDMON *et al.*, 2016). A figura 14 mostra as etapas de classificação realizados pela rede *YOLO*.

As limitações da primeira versão da rede *YOLO* são: dificuldade de localizar objetos pequenos e próximos, problemas para generalizar distorções e localizações incorretas. Porém, novas versões da rede foram desenvolvidas com o objetivo de resolver tais limitações. O presente trabalho utiliza a quarta versão da rede, em que diversas otimizações foram realizadas para melhorar o desempenho e a velocidade, além das melhorias para detectar pequenos objetos (REDMON *et al.*, 2016; JIANG *et al.*, 2022). Além disso, diversas técnicas de aumento de dados foram utilizadas para melhorar a capacidade de generalização da rede. Tais técnicas consistem em combinar, distorcer ou alterar as cores de uma imagens

[3] Segundo o artigo original, os testes foram conduzidos utilizando uma placar gráfica NVIDIA Titan X.

Figura 14 – Exemplo de funcionamento de uma rede *YOLO*: a) imagem original contendo os objetos a serem detectados; b) cada célula da grade representa classificação do objeto; c) *bounding boxes* candidatos; d) cruzamento das infomações de (b) e (c).



Fonte – Daniel Freire Tsuha, 2023

para aumentar a quantidade de dados do conjunto de treinamento (GAO; CAI; MING, 2020).

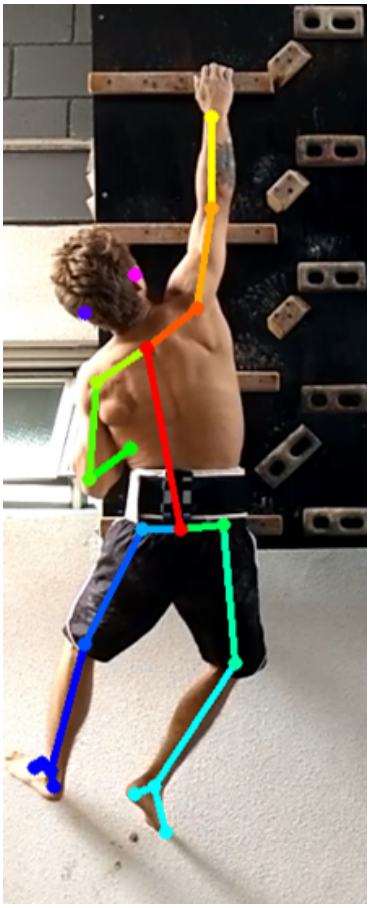
2.5.2 OpenPose

A rede *OpenPose* foi apresentada em Cao *et al.* (2019) e tem por objetivo estimar a pose de humanos em vídeos. O diferencial dessa implementação é a abordagem *bottom-up*, que detecta regiões do corpo humano e a direção dos membros para estimar a pose. Outras abordagens utilizam técnicas *top-down*, em que primeiro se detecta as pessoas na cena e depois estima-se as poses. Em comparação, as abordagens *bottom-up* tendem a ser mais resilientes quando há pessoas próximas na imagem, já que os detectores de pessoas da outra abordagem podem apresentar maiores dificuldades nesses casos. Além disso, o tempo de execução não é acoplado ao número de pessoas, já que não é necessário executar um estimador de pose para cada pessoa (CAO *et al.*, 2019).

Para estimar a pose das pessoas na cena, a rede *OpenPose* realiza duas etapas de processamento: a predição dos campos de afinidade das partes (Part Affinity Fields - PAF) e os mapas de confiança. Os campos de afinidade são vetores bidimensionais que indicam a localização e a orientação dos membros. Já os mapas de confiança indicam a probabilidade de uma determinada região conter uma parte do corpo. Após a localização dos membros utilizando os mapas de confiança, o algoritmo cria uma representação de regiões conectadas (i.e. grafo) com tais pontos (CAO *et al.*, 2019).

Após a criação do grafo, a estimativa de pose se torna um problema de partição de grafos, já que o objetivo é descobrir quais pontos pertencem a quais pessoas. Após a redução do número de arestas utilizando diversas técnicas de relaxamento, os campos de afinidade são combinados com os resultado obtidos até o momento para aumentar a confiabilidade e a precisão da rede (CAO *et al.*, 2019). No presente trabalho, a rede *OpenPose* foi utilizada para estimar o centro de massa dos escaladores e a figura 15 mostra um exemplo de esqueletização da rede.

Figura 15 – Exemplo de esqueletização de um atleta utilizando o OpenPose.



Fonte – Daniel Freire Tsuha, 2023

2.6 Conclusão

Neste capítulo foi apresentado o ferramental teórico utilizado no desenvolvimento da pesquisa. Foram detalhadas as redes neurais convolucionais adaptadas para detecção das agarras e dos atletas. Tais redes foram contextualizadas em tópicos mais amplos, como redes neurais e aprendizagem de máquina. Além disso foi descrito como computadores digitais representam imagens e os aspectos técnicos da prova de escalada esportiva de velocidade.

3 Revisão bibliográfica

Técnicas de visão computacional são algoritmos utilizados para extrair informações de imagens e vídeos (JAIN, 1989; SZELISKI, 2010). Tais algoritmos têm aplicações práticas em diversas áreas do conhecimento, auxiliando na tomada de decisão e na automação de processos. Na medicina, esses algoritmos podem auxiliar no diagnóstico de doenças com base em testes de imagem (RIBEIRO; NUNES, 2022); na engenharia, os carros usam essas técnicas para dirigir de forma autônoma (YENIKAYA; YENIKAYA; DUVEN, 2013); em sistemas de segurança, algoritmos de visão computacional podem ser utilizados para reconhecer pessoas e identificar comportamentos suspeitos (VEZZANI; BALTIERI; CUCCHIARA, 2013). Outra área de conhecimento que pode se beneficiar das aplicações de tais técnicas é a ciência do esporte.

As gravações de competições esportivas permitem que técnicos e treinadores analisem detalhadamente o desempenho de atletas e equipes. A avaliação pode analisar aspectos como: a quantidade de eventos (e.g. número de faltas e escanteios durante uma partida de futebol); a execução de um movimento (e.g. a postura de um ginasta ao realizar um salto); ou a capacidade física de um atleta (e.g. a velocidade e a aceleração de um corredor).

A detecção automática dos movimentos de atletas de diferentes esportes tem recebido considerável atenção da comunidade científica nos últimos anos. Tais sistemas têm o potencial de entender melhor os atletas e adaptar o treinamento ao potencial de cada um. No entanto, o desenvolvimento de um sistema confiável com bons níveis de precisão e baixo custo continua sendo um problema em aberto. De acordo com Barris e Button (2008), os atletas tendem a realizar movimentos rápidos, sujeitos a mudanças imprevisíveis de direção e, até, colisões com outros participantes, e essas condições violam os pressupostos de movimentos suaves nos quais os algoritmos de visão computacional são tipicamente baseados.

Uma maneira de avaliar os atletas é marcar manualmente os pontos de interesse usando *softwares* como *Kinovea* (CHARMANT, 2017). No entanto, esse tipo de análise consome tempo e depende de mão de obra especializada. O uso de marcadores refletivos, como o sistema *Vicon*[1], é considerado o padrão-ouro para a realização de análises biomecânicas. No entanto, esse tipo de equipamento não pode ser utilizado em partidas oficiais, além de ter um alto custo. Outros sensores específicos, como acelerômetros e

[1] ⟨https://www.vicon.com/⟩

sensores de pressão, também não podem ser usados em partidas oficiais. Além disso, o uso desses dispositivos pode prejudicar o desempenho dos atletas.

Nesse contexto, algoritmos de visão computacional podem ser utilizados como alternativa para avaliar de forma automática o desempenho de atletas durante a prática esportiva. Sendo assim, o objetivo da presente revisão é levantar as principais técnicas computacionais utilizadas para avaliar aspectos físicos e técnicos de atletas. Também foram levantadas informações sobre frequência e resolução das imagens utilizadas, métricas analisadas e formas de avaliação.

O presente capítulo está organizado da seguinte forma: a seção 3.1 apresenta os trabalhos relacionados; a seção 3.2 apresenta o protocolo de pesquisa; a seção 3.3 apresenta os esportes e as métricas estudadas; a seção 3.4 apresenta as principais técnicas computacionais utilizadas para processar os vídeos; a seção 3.5 apresenta as amostras de atletas e os conjuntos de dados utilizados; a seção 3.6 apresenta como os algoritmos são avaliados; a seção 3.7 mostra os próximos passos das pesquisas analisadas; a seção 3.8 apresenta as tendências da área; e, por fim, a seção 3.9 apresenta as conclusões da revisão.

3.1 Trabalhos relacionados

Durante a etapa de revisão bibliográfica foram encontradas outras revisões correlatas. Esta seção apresenta os objetivos e enfoques de cada uma delas.

Em Barris e Button (2008), os autores apresentam as ferramentas de rastreamento existentes aplicadas à vigilância e aos esportes. Também são apresentadas as principais ferramentas comerciais utilizadas no âmbito esportivo, assim como as limitações de cada uma.

Thomas *et al.* (2017) explicam como as técnicas de visão computacionais têm sido aplicadas ao esporte. O artigo apresenta algumas tarefas, tais como o rastreamento de jogador e bola, a análise de movimento, a rotulação automática de eventos e a aplicação em sistemas comerciais. Os autores também discutem os impactos dessas tecnologias e apresentam diversos conjuntos de dados de imagens esportivas.

A revisão de Shih (2018) apresenta técnicas para o reconhecimento, a compreensão e a organização de conteúdo. As tarefas descritas no artigo incluem detecção de objetos, reconhecimento de ações e ventos, inferência contextual e análise semântica.

O artigo de Colyer *et al.* (2018) tem por objetivo apresentar as abordagens sem marcadores para realização de análises biomecânicas e cinemáticas. As técnicas apresentadas são: reconstrução do movimento utilizando um modelo tridimensional, modelos generativos e abordagem discriminativa. Além disso, os autores apresentam uma análise da precisão das técnicas.

Também foram encontradas três revisões voltadas especificamente para o futebol. Em Al-Ali e Almaadeed (2017), os autores apresentam técnicas de rastreamento baseadas em extração de características. Manafifard, Ebadi e Moghaddam (2017) também analisam ferramentas de rastreamento, porém com o foco em categorizar as ferramentas existentes e apresentar os pontos fortes e fracos de cada uma. Por fim, o artigo de Fischer, Keim e Stein (2019) apresenta as principais tarefas encontradas na literatura, assim como uma comparação entre as ferramentas desenvolvidas em contextos comerciais e acadêmicos. Os autores também avaliam os artigos por nível de complexidade da tarefa realizada.

Apesar de não ser uma revisão focada em processamento de imagem e visão computacional, o artigo de Kruk e Reijne (2018) compara as principais ferramentas de análise de movimento utilizados nos esportes. Os autores classificam as ferramentas utilizando critérios como área de captura e precisão e apresentam os pontos positivos e negativos de cada abordagem.

Diferentemente dos trabalhos citados, esta revisão tem por objetivo levantar as principais técnicas de processamento de imagem e visão computacional utilizadas para avaliar aspectos físicos ou técnicos de atletas a partir de vídeos e focando em aspectos individuais. Os artigos analisados apresentam exemplos de imagens processadas, além das etapas de processamento, extração e avaliação.

3.2 Protocolo

A revisão foi conduzida seguindo as etapas de planejamento, execução e sumarização (KITCHENHAM, 2004). O objetivo da pesquisa foi identificar estudos que utilizem técnicas de visão computacional para avaliar o desempenho de atletas sem a utilização de marcadores ou equipamentos específicos. As questões que a pesquisa se propõe a responder são:

- quais técnicas de visão computacional são utilizadas para avaliar o desempenho de atletas durante a prática esportiva?

- quais métricas são extraídas a partir de vídeos de prática esportiva?
- quais são as formas de avaliação das ferramentas desenvolvidas?

Para que um artigo fosse selecionado, ele deveria satisfazer os seguintes critérios:

- aplicar técnicas de visão computacional ou processamento de imagens à vídeos de prática ou treinamento esportivo;
- mensurar aspectos individuais físicos ou técnicos de atletas durante a prática esportiva;
- apresentar a forma de avaliação e validação do algoritmo desenvolvido.

Para que um artigo não fosse considerado, ele deveria atender à um dos seguintes critérios:

- tarefas de processamento de imagem (como rastreamento ou reconstrução de movimento de atletas) sem extração de métricas ou análise de desempenho;
- análise de equipes;
- utilização de qualquer tipo de marcador ou câmeras equipadas com sensores de profundidade;
- utilização exclusiva de equipamentos específicos (Acelerômetro, GPS, plataforma de pressão etc.);
- Revisões sistemáticas.

As palavras-chave utilizadas durante a busca foram separadas em duas categorias: computação e esportes. As palavras de ambas as categorias foram combinadas em pares e conectadas com o operador lógico *AND* para formar as *strings* de busca. Dependendo do funcionamento da base de dados foram consideradas as variações de uma palavra (como *sport* e *sports*). Foi necessária a utilização dessa abordagem, pois algumas máquinas de busca limitam o número de palavras-chave. A tabela 1 mostra as *strings* utilizadas. Foram considerados os artigos publicados entre 2008 e 2020.

A busca foi realizada em 10/01/2020 nas bases de dados *IEEE Xplorer*[2], *ACM Digital Library*[3], *PubMed*[4], *ScienceDirect*[5] e *Springer Link*[6]. As *strings* de busca *(bi-*

[2] ⟨https://ieeexplore.ieee.org⟩
[3] ⟨https://dl.acm.org⟩
[4] ⟨https://www.ncbi.nlm.nih.gov/pubmed⟩
[5] ⟨https://www.sciencedirect.com⟩
[6] ⟨https://link.springer.com/⟩

Tabela 1 – Palavras-chave combinadas em pares para montar as *strings* de busca

Visão computacional	**Esporte**
computer vision	*sport*
vision based	*biomechanic*
video based	*motion analysis*
video analysis	*player*
markless	
image processing	

Fonte – Daniel Freire Tsuha, 2023

omechanic AND "image processing")*, *("motion analysis" AND "computer vision")* e *("motion analysis" AND "image processing")*, quando submetidas à máquina de busca do *IEEE Xplorer*, retornaram mais de mil resultados cada. Para limitar os resultados e viabilizar a pesquisa, foi adicionada a expressão *AND "sport"* ao final da *string*.

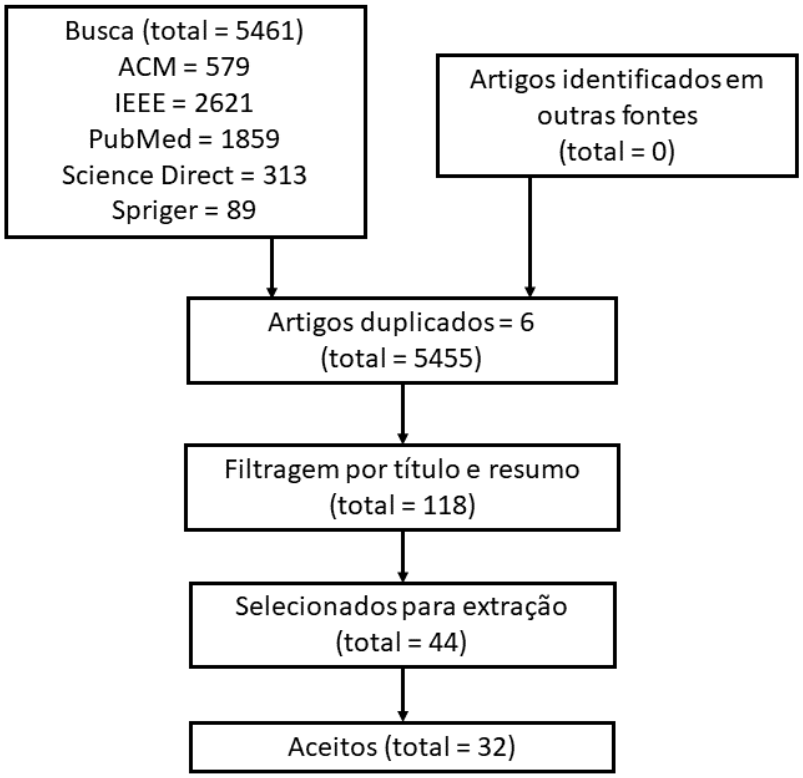
A busca resultou em um total de 5461 artigos dos quais 118 foram selecionados com base no título e resumo. Após a aplicação dos critérios de inclusão e exclusão, 44 artigos seguiram para a fase de extração, sendo que 32 satisfaziam as condições necessárias. Dentre esses artigos, dois não eram aplicados à prática esportiva, porém as tarefas realizadas são atividades que demandam esforço físico e as metodologias utilizadas são relevantes para a pesquisa. A figura 16 mostra o processo de escolha dos artigos.

O interesse em extrair informações de imagens esportivas apresenta uma tendência crescente. Entre 2008 e 2016, houve pouca variação no número de artigos publicados. Nos anos de 2017 e 2018, houve um aumento no interesse pelo assunto. Apesar de uma pequena queda em 2019, o número de artigos no ano ficou acima da média dos anos anteriores. A figura 17 mostra o número de artigos por ano.

3.3 Esportes e métricas

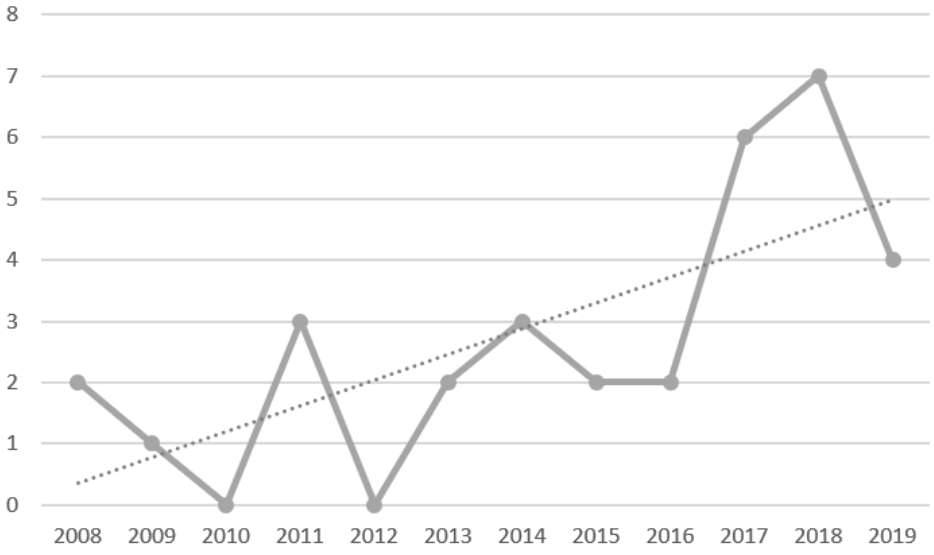
Foram extraídos dos artigos os exercícios estudados e as métricas analisadas. No total, 21 atividades foram analisadas: sendo treze esportes, cinco atividades físicas e três modalidades de ginástica artística. O esporte mais estudado foi a corrida, sendo contemplada por seis artigos (18,5%). Já a caminhada foi contemplada em quatro artigos (12,5%), sendo a segunda atividade mais estudada. A figura 18 mostra a quantidade de artigos por categoria de esporte. Além das métricas relacionadas aos passos (número, frequência, duração, comprimento médio, ritmo, etc.), Gade, Larsen e Moeslund (2017) e

Figura 16 – Etapas da revisão sistemática.



Fonte – Daniel Freire Tsuha, 2020

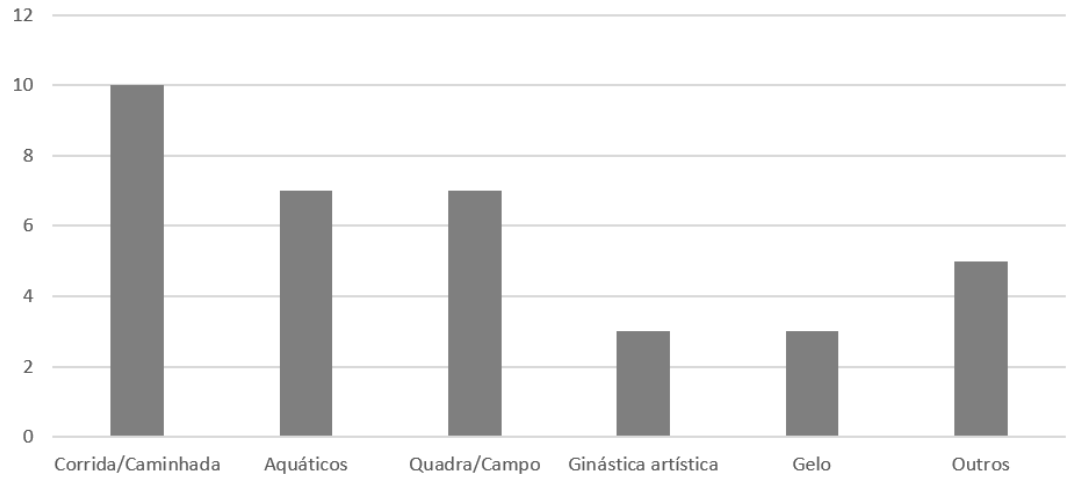
Figura 17 – Artigos publicados por ano dentre os selecionados.



Fonte – Daniel Freire Tsuha, 2020

Koporec *et al.* (2018) mensuraram o gasto energético dos atletas. Além disso, Koporec *et al.* (2018) também estimaram a frequência cardíaca durante a prática de caminhada. A tabela 2 mostra as métricas de corrida e caminhadas estudas.

Figura 18 – Artigos por categoria de esporte.



Fonte – Daniel Freire Tsuha, 2020

Tabela 2 – Corrida e caminhada.

Esporte	**Métrica**	**Artigo**
Corrida	Análise das articulações	El-sallam *et al.* (2013)
	Velocidade	El-sallam *et al.* (2013) Nagano *et al.* (2017)
	Velocidade de transição	Yagi *et al.* (2018)
	Ritmo de corrida	Jinyan, Guanlei e Yu (2013)
	Frequência de passos	Evans *et al.* (2018)
	Número de passos	Yagi *et al.* (2018)
	Duração dos passos	Yagi *et al.* (2018) Evans *et al.* (2018)
	Comprimento médio da passada	Yagi *et al.* (2018)
	Gasto energético	Gade, Larsen e Moeslund (2017)
Caminhada	Análise das articulações	Ong, Harris e Hamill (2017) El-sallam *et al.* (2013)
	Velocidade	El-sallam *et al.* (2013)
	Comprimento médio da passada	Barone *et al.* (2016)
	Frequência cardíaca	Koporec *et al.* (2018)
	Gasto energético	Koporec *et al.* (2018)

Fonte – Daniel Freire Tsuha, 2020

Entre os esportes aquáticos, a natação foi objeto de estudo de quatro artigos (12,5%), juntamente com a caminhada, foi o segundo esporte mais estudado. Dado à natureza do esporte, as métricas focam em aspectos da braçada durante a prática do esporte (distância,

intervalo e número de braçadas). Já em relação ao salto ornamental, os estudos focam em aspectos técnicos da execução do exercício como pontuação (PARMAR; MORRIS, 2017) e a sincronia entre os atletas (DING *et al.*, 2008). Cronin *et al.* (2019) estuda a corrida em águas profundas, uma atividade que consiste realizar o movimento de corrida dentro de uma piscina sem tocar os pés no chão. Nesse artigo, o objetivo foi mensurar a distância dos passos e realizar a análise das articulações. A tabela 3 detalha as métricas extraídas de cada atividade.

Tabela 3 – Esportes aquáticos.

Esporte/Atividade	**Métrica**	**Artigo**
Natação	Análise das articulações	Ceseracciu *et al.* (2011)
	Distância por braçada	Sha *et al.* (2014)
	Intervalos cíclicos	Zecha e Lienhart (2015)
	Velocidade	Sha *et al.* (2014) Ceseracciu *et al.* (2011)
	Número de braçadas	Sha *et al.* (2014) Victor *et al.* (2017)
ornamental	Pontuação	Parmar e Morris (2017)
	Sincronia	Ding *et al.* (2008)
Corrida em águas profundas	Análise das articulações	Cronin *et al.* (2019)
	Duração do passo	Cronin *et al.* (2019)

Fonte – Daniel Freire Tsuha, 2020

Em relação aos esportes de quadra e campo, o futebol foi a atividade mais analisada (3 artigos, 9,3%). Apesar de ser um esporte praticado em equipes, foram considerados estudos que extraiam métricas de forma individual. Além de métricas de desempenho físico, como distância percorrida e velocidade (WU *et al.*, 2019), Leo, D'Orazio e Trivedi (2009) mensurou o nível de participação dos atletas durante a partida e Sato *et al.* (2015) avaliou os chutes de estudantes com diferentes níveis de habilidade.

O tênis foi estudado por dois dos artigos encontrados. O nível de habilidade dos atletas foi mensurado por Mukai, Asano e Hara (2011), já Sheets *et al.* (2011) analisou o movimento das articulações e mediu a velocidade dos atletas. Koporec *et al.* (2018) mensuraram a frequência cardíaca e o gasto energético de jogadores de squash. Por fim, Chu e Situmeang (2017) classificou as estratégias utilizada por jogadores de badminton. A tabela 4 apresenta os detalhes dos esportes de quadra e campo encontrados na literatura.

Outras categorias de esportes foram encontradas com menor frequência. A busca retornou três artigos (9,3%) que analisaram diferentes modalidades dentro da ginástica artística. Em todos os casos, o objetivo dos autores foi pontuar de forma automática a

Tabela 4 – Esportes de quadra e campo.

Esporte	**Métrica**	**Artigo**
Futebol	Distância percorrida	Wu *et al.* (2019)
	Velocidade	Wu *et al.* (2019)
	Nível de participação	Leo, D'Orazio e Trivedi (2009)
	Avaliação de chute	Sato *et al.* (2015)
Tênis	Nível de habilidade	Mukai, Asano e Hara (2011)
	Análise das articulações	Sheets *et al.* (2011)
	Velocidade	Sheets *et al.* (2011)
Squash	Frequência cardíaca	Koporec *et al.* (2018)
	Gasto energético	Koporec *et al.* (2018)
Badminton	Análise tática	Chu e Situmeang (2017)

Fonte – Daniel Freire Tsuha, 2020

execução dos exercícios. Também foram encontrados três artigos (9,3%) que analisaram esportes praticados no gelo. Assim como na ginástica artística, Parmar e Morris (2017) também pontuaram os movimentos de patinação artística. Os detalhes dos artigos que analisam a ginastica artística e os esportes praticados no gelo podem ser encontrados nas tabelas 5 e 6 respectivamente.

Tabela 5 – Ginástica artística.

Modalidade	**Métrica**	**Artigo**
Salto sobre a mesa	Pontuação	Parmar e Morris (2017)
Ginástica Rítmica	Pontuação	Díaz-Pereira *et al.* (2014)
Barra horizontal	Pontuação	Shin e Ozawa (2008)

Fonte – Daniel Freire Tsuha, 2020

Tabela 6 – Esportes no gelo.

Esporte	**Métrica**	**Artigo**
Salto com esqui	Força/Pico de força	Zecha *et al.* (2018)
Esqui estilo livre	Nível de habilidade	Wang *et al.* (2019)
Patinação artística	Pontuação	Parmar e Morris (2017)

Fonte – Daniel Freire Tsuha, 2020

Além de esportes como ciclismo e salto em distância, outras atividades foram consideradas relevantes. Em Boonim e Sanguansat (2018), os autores apresentaram um sistema para cronometrar a execução de um exercício realizado com uma escada. Mehrizi *et al.* (2017), Mehrizi *et al.* (2018) analisam as articulações e o momento durante o levantamento de uma caixa em diversas alturas. Já Yu *et al.* (2019) mensuram a fadiga de trabalhadores da construção civil. Apesar desses artigos não serem referentes à esportes

ou atividades físicas, as métricas analisadas e as metodologias utilizadas são relevantes para a presente análise. A tabela 7 exibe a lista completa de atividades e métricas.

Tabela 7 – Outros esportes e atividades.

Esporte/Atividade	**Métrica**	**Artigo**
Ciclismo	Análise das articulações	Gastel *et al.* (2014)
	Frequência cardíaca	Gastel *et al.* (2014)
Salto em distância	Análise das articulações	El-sallam *et al.* (2013)
Treino em escada	Tempo de execução	Boonim e Sanguansat (2018)
Levantamento de caixa	Análise das articulações	Mehrizi *et al.* (2018)
	Momento (Física)	Mehrizi *et al.* (2017)
Construção civil	Fadiga	Yu *et al.* (2019)

Fonte – Daniel Freire Tsuha, 2020

Mesmo cada atividade possuindo suas peculiaridades, algumas métricas são comuns à vários esportes. A análise de articulações foi a métrica mais estudada (7 artigos, 21,8%), sendo uma variável importante em diversos esportes. Outras métricas como velocidade (citado por 6 artigos, 18,7%) e pontuação (citado por 4 artigos, 12,5%) se mostraram relevantes no contexto esportivo.

3.4 Técnicas computacionais

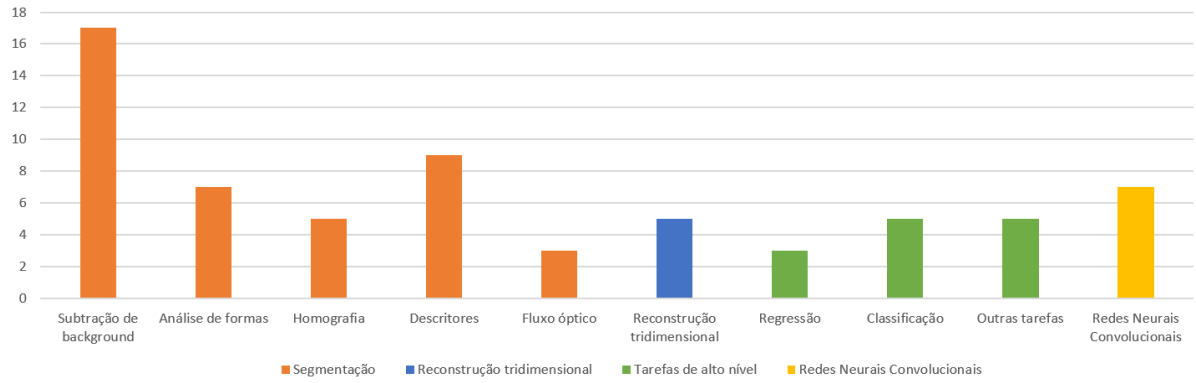
Esta seção apresenta as principais técnicas computacionais utilizadas pelos estudos analisados. Primeiramente são apresentados os algoritmos de processamento de imagem (Subseção 3.4.1). A subseção 3.4.2 lista as tarefas de alto nível e, por fim, a subseção 3.4.3 apresenta os algoritmos de redes neurais convolucionais utilizadas por alguns dos estudos. A figura 19 mostra a frequência das técnicas computacionais utilizadas.

3.4.1 Processamento de imagens

As tarefas de processamento de imagem encontradas durante a revisão foram divididas em segmentação, extração de características e reconstrução tridimensional do movimento.

Devido às particularidades de cada esporte, não foi possível identificar um processo comum entre os estudos. Sendo assim, esta seção apresenta os algoritmos mais relevantes utilizados para o processamento de imagens esportivas.

Figura 19 – Técnicas computacionais.



Fonte – Daniel Freire Tsuha, 2020

Segmentação

Segundo Gonzalez e Woods (2011), a segmentação consiste em separar as áreas de uma imagem de forma a identificar as regiões de interesse. No contexto esportivo, a utilização dessas técnicas permite detectar, de forma automática, jogadores, linhas de marcação e objetos. Dependendo do ambiente em que o esporte é praticado, certas abordagens apresentam melhores resultados.

Subtração de background: umas das formas mais comuns de segmentar uma região de interesse é a utilização de subtração de *background*, sendo utilizada por 17 dos estudos analisados (53,1%). Essa técnica calcula a diferença entre uma imagem contendo os objetos de interesse e uma referência contendo apenas o cenário de fundo. A abordagem mais simples consiste em capturar, previamente, uma imagem de referência e utilizá-la para subtrair os quadros do vídeo. Essa abordagem foi utilizada por Shin e Ozawa (2008), Gastel *et al.* (2014), Gade, Larsen e Moeslund (2017), Leo, D'Orazio e Trivedi (2009), Nagano *et al.* (2017) e El-sallam *et al.* (2013). Mukai, Asano e Hara (2011), Victor *et al.* (2017) e Sheets *et al.* (2011) não especificaram o algoritmo utilizado.

Quando uma imagem de referência não está disponível, é possível utilizar técnicas mais sofisticadas para reconstruir o *background*. Em Ding *et al.* (2008), a diferença entre dois quadros consecutivos é utilizada para estimar o movimento da câmera e reconstruir o *background*. Chu e Situmeang (2017), Leo, D'Orazio e Trivedi (2009), Wang *et al.* (2019), Boonim e Sanguansat (2018) e Ceseracciu *et al.* (2011) utilizam um modelo de mistura gaussiana para estimar os valores dos píxeis de fundo ao longo do tempo. Esse algoritmo

é conhecido como *Mixture of Gaussian (MoG)* ou *Gaussian Mixture Model (GMM)*. Já Evans *et al.* (2018) utilizam a técnica *Independent Multimodal Background Subtraction* (IMBS) para reconstruir o *background* e segmentar as regiões de interesse. El-sallam *et al.* (2013) utilizaram o algoritmo *Kernel Density Estimation* (KDE) para reconstruir o modelo do *background*.

Análise de formas: analisar o formato de objetos também pode ser uma maneira de encontrar regiões de interesse. Utilizando as cores dos *pixels* ou a saída de um algoritmo de subtração de *background*, diversos algoritmos podem ser utilizados para detectar determinados objetos ou melhorar os resultados de etapas anteriores. Sete artigos utilizam tais técnicas para detectar regiões de interesse (21,8%).

Em Chu e Situmeang (2017), os autores utilizam a transformada de Hough probabilística para identificar as linhas da quadra (MATAS; GALAMBOS; KITTLER, 2000). Tal técnica consiste em mapear todas as possíveis retas de uma imagem para outro espaço matemático de forma eficiente (GONZALEZ; WOODS, 2011). Já em Yagi *et al.* (2018) e Sha *et al.* (2014), os autores utilizaram o algoritmo *Random Sample Consensus* (RANSAC) para detectar as linhas. Essa abordagem foi utilizada pois permite detectar retas mesmo na presença de ruído (FISCHLER; BOLLES, 1981).

Em outros estudos, após a etapa de subtração de *background*, as formas das regiões encontradas são analisadas para melhorar o resultado da segmentação. Gade, Larsen e Moeslund (2017) utilizam a proporção da região de interesse para eliminar a sombra do atleta. Já Ding *et al.* (2008), subdivide a região segmentada de forma a facilitar a análise de atletas de salto sincronizado.

Homografia: devido ao posicionamento da câmera em relação à área em que a prática esportiva é realizada, as imagens capturadas possuem distorções, o que impede o mapeamento direto da posição do atleta para a posição real dentro da área de prática. Para contornar esse problema, cinco artigos (15,6%) utilizam técnicas de homografia para corrigir a distorção das imagens. Chu e Situmeang (2017) utilizam tal técnica para mapear a posição dos jogadores em quadra. Em Yagi *et al.* (2018), a homografia é utilizada para estimar a posição dos corredores em uma pista de cem metros. Já em Sha *et al.* (2014), a técnica é utilizada para corrigir o movimento da câmera em relação à piscina. Barone *et al.* (2016) utilizam esse algoritmo para mapear a posição do atleta em relação à esteira. Por fim, Boonim e Sanguansat (2018) aplica o algoritmo para mapear a posição dos pés do atleta em relação a uma escada e identificar o início e o fim do exercício.

Extração de características

A extração de características é uma etapa intermediária que normalmente é realizada depois da segmentação. Após a identificação das regiões de interesse, esses algoritmos são capazes de extrair diversos aspectos de tais áreas. No caso da extração de características de vídeos, alguns algoritmos não necessitam da etapa de segmentação, já que a análise é feita por meio da comparação de dois quadros consecutivos. Os dados gerados nessa etapa do processamento podem ser utilizados como entrada para algoritmos de reconhecimento de padrões ou de classificação para analisar os movimentos dos atletas.

Descritor de características: esses algoritmos são utilizados para descrever as regiões segmentadas, levando em conta características como área, morfologia, fronteira, textura etc. (GONZALEZ; WOODS, 2011). Nove dos artigos encontrados utilizavam algum algoritmo desse tipo (28,1%).

O algoritmo mais utilizado foi o *Histogram of Oriented Gradients* (HOG) (CHU; SITUMEANG, 2017; SATO *et al.*, 2015; SHA *et al.*, 2014; ZECHA; LIENHART, 2015; MEHRIZI *et al.*, 2018). Nesse algoritmo, a imagem é dividida em células e o gradiente de cada região é calculado, gerando, assim, o histograma correspondente. O gradiente de uma imagem indica a direção em que há a maior variação entre píxeis vizinhos (GONZALEZ; WOODS, 2011; SUARD *et al.*, 2006)

Em Sha *et al.* (2014), o algoritmo utilizado para descrever as regiões segmentadas foi o *Scale-Invariant Feature Transform* (SIFT). Esse algoritmo extrai vetores de características locais que são invariantes à rotação, redimensionamento e, parcialmente, à iluminação. O algoritmo possui inspiração biológica e se baseia em como o cérebro processa imagens (LOWE, 1999).

Outra característica que pode ser extraída de uma imagem é o centroide geométrico de uma região de interesse. Esse ponto representa o centro de gravidade de uma área (WEISSTEIN, 2020) e pode ser utilizado como referência para calcular a trajetória ou velocidade de um elemento. Ding *et al.* (2008), Barone *et al.* (2016) e Nagano *et al.* (2017) utilizaram essa abordagem.

Fluxo óptico: é uma classe de algoritmos utilizados para estimar a velocidade aparente de superfícies em vídeos. A análise desses dados pode ajudar a diferenciar objetos e detectar padrões de movimento (HORN; SCHUNCK, 1981; BEAUCHEMIN; BARRON,

1995; WANG *et al.*, 2011). Apesar de apenas três estudos utilizarem essa técnica (9,3%), seis algoritmos diferentes foram utilizados.

Em Sato *et al.* (2015), os autores analisam diversos descritores para estimar o nível de habilidade de jogadores ao chutar uma bola. Para alcançar tal objetivo, primeiramente é extraída a trajetória densa das imagens, esse algoritmo utiliza diversas escalas espaciais para calculá-la (WANG *et al.*, 2011). A próxima etapa é calcular o *HOG*, o *Histogram of Optical Flow* (HOF) e o *Motion Boundary Histogram* (MBH) utilizando as trajetórias. Esses dados são processados para estimar a qualidade do chute realizado.

Koporec *et al.* (2018) utiliza o algoritmo desenvolvido por Farneback (2003) para calcular o fluxo óptico denso. Com a saída desse algoritmo, são utilizados os descritores *Histograms of Oriented Optical Flow* (HOOF) e *Histograms of Absolute Flow Amplitudes* (HAAF) para estimar o gasto calórico do atleta. Segundo Chaudhry *et al.* (2009), o HOOF é um descritor não euclidiano que independe da escala e da direção de movimento. Já o descritor HAAF mensura a amplitude do movimento a partir do histograma de fluxo óptico (PERS *et al.*, 2010).

Já em Cronin *et al.* (2019), o objetivo do estudo é reconhecer e pontuar movimentos de ginástica artística. Para tanto, os autores utilizam o algoritmo de Farneback (2003) para calcular o fluxo óptico e o descritor *Motion Vector Flow Instance* (MVFI) para codificar informações de velocidade (CRONIN *et al.*, 2019).

Reconstrução tridimensional

A abordagem de reconstrução tridimensional do movimento se baseia na utilização de diversas câmeras para capturar a mesma cena de diversos ângulos. A partir das diferentes imagens, esses algoritmos são capazes de extrair informações de profundidade, o que permite a reconstrução do movimento e a extração de várias métricas. Cinco artigos utilizam essa abordagem (15,6%).

Apesar dos diferentes contextos em que essa técnica foi empregada, foi possível identificar algumas etapas em comum entre os estudos analisados:

1. Captura de um modelo do corpo do atleta em posição estática;
2. Captura do movimento utilizando diversas câmeras;
3. Reconstrução do movimento;

4. Encaixe do modelo estático com o modelo de movimento para mapear as regiões de interesse;
5. Extração das métricas de desempenho.

A captura de um modelo estático do corpo do atleta pode ser feita de diversas formas e ter diversas finalidades. Em El-sallam *et al.* (2013), uma imagem DEXA (*Dual-Energy X-ray Absorptiometry*) do atleta é utilizada para estimar o centro de massa do corpo durante o movimento. O equipamento utilizado para gerar essas imagens é capaz de medir a densidade de cada região do corpo. Sheets *et al.* (2011) utilizam um escâner tridimensional para criar um modelo do atleta que é, então, subdividido em quinze partes para identificação dos membros após o encaixe. No estudo de Ceseracciu *et al.* (2011), os autores utilizaram várias câmeras para extrair a silhueta do atleta e recriar o modelo tridimensional, esse modelo é subdividido para auxiliar no reconhecimento das articulações do atleta.

Diversas abordagens podem ser utilizadas para reconstruir o movimento. Em El-sallam *et al.* (2013), Sheets *et al.* (2011) e Ceseracciu *et al.* (2011), os autores utilizaram a subtração de *background* para extrair a silhueta das imagens capturadas por diversos ângulos. As silhuetas são combinadas de forma a reconstruir o volume do corpo do atleta, o resultado dessa operação é chamado de *visual hull*.

Já em Mehrizi *et al.* (2017) e Mehrizi *et al.* (2018), os autores utilizaram uma variação do algoritmo *Twin Gaussian Process* (TGP) para reconstruir o movimento. Esse algoritmo é capaz de reconstruir o movimento humano em três dimensões sem a necessidade de calibração da câmera ou inicialização da pose inicial (BO; SMINCHISESCU, 2009). A utilização dessa abordagem dispensa o modelo estático do corpo do atleta e, consequentemente, a etapa de encaixe dos modelos.

Para encaixar os modelos de movimento com o modelo estático, El-sallam *et al.* (2013) apresentam duas abordagens: uma mais rápida e menos precisa e outra lenta e mais precisa. Na primeira abordagem, o centroide do *visual hull* é utilizado como referência para alinhar com o modelo estático, então uma aproximação linear é utilizada para realizar o ajuste fino e distribuir os pontos da malha; a segunda abordagem ajusta o modelo estático do atleta à uma estimativa de esqueleto calculada a partir do *visual hull*. Sheets *et al.* (2011) utilizam o algoritmo *Iterative Closest Point* (ICP) juntamente com o algoritmo de *Levenberg–Marquardt* (CORAZZA *et al.*, 2009). Já em Ceseracciu *et al.* (2011), os autores

utilizam o algoritmo proposto em Corazza *et al.* (2010), que tem por objetivo encontrar os parâmetros de rotação que minimize a distância entre os modelos.

3.4.2 Reconhecimento de padrões, classificação e regressão

Os dados obtidos após a etapa de processamento e extração de características das imagens podem ser usados em tarefas como identificação ou avaliação de um movimento. Para tanto, algoritmos estatísticos e de aprendizagem de máquina são utilizados. Treze dos artigos revisados fazem o uso dessa abordagem (46,8%).

Classificação

Algoritmos de classificação são utilizados para identificar a qual classe pertence um elemento dado os seus atributos. No contexto esportivo, essa abordagem permite identificar o tipo do movimento realizado por um atleta ou ajudar na avaliação de um determinado movimento.

Em Zecha e Lienhart (2015), os autores utilizam uma *Support Vector Machine* (SVM) para identificar o tipo de movimento realizado por jogadores de badminton. Já em Chu e Situmeang (2017), uma SVM foi treinada para identificar as etapas do movimento durante a prática de natação. No trabalho de Wang *et al.* (2019), esse algoritmo foi utilizado para classificar a pose do esquiador.

Ainda em Chu e Situmeang (2017), os autores utilizam o classificador *Hidden Markov Model* (HMM) para classificar a estratégia do jogador (ofensiva/defensiva). Em Leo, D'Orazio e Trivedi (2009), esse algoritmo é utilizado para estimar o nível de participação de uma atleta durante partidas de futebol.

Em Mehrizi *et al.* (2017) e Mehrizi *et al.* (2018) os autores utilizam o *Twin Gaussian Processes* (TGP) para identificar as articulações do corpo e reconstruir o esqueleto tridimensional de uma pessoa levantando uma caixa.

Análise de regressão

A análise de regressão pode ser aplicada quando se deseja estimar o valor de uma variável utilizando como base a correlação entre os dados. Mukai, Asano e Hara (2011) avaliam o desempenho de jogadores de tênis utilizando métricas como velocidade e altura do arremesso da bola. Essas métricas são então utilizadas para criar um modelo de regressão múltipla tendo como base a avaliação humana para descobrir os valores dos parâmetros.

Com o objetivo de atribuir notas de forma automática à movimentos realizados por atletas, Parmar e Morris (2017) utilizaram uma *Support Vector Regression* (SVR) combinada com outros algoritmos. Em um dos experimentos apresentados em Koporec *et al.* (2018), esse algoritmo também foi aplicado em uma das etapas do processo para estimar o gasto energético dos atletas.

Outros algoritmos

Apesar de ser um algoritmo de classificação, o *Linear Discriminant Analysis* (LDA) foi utilizado por Díaz-Pereira *et al.* (2014) juntamente com o *Principal Component Analysis* (PCA) para reduzir a dimensionalidade dos dados. Nesse artigo, o algoritmo *K-Nearest Neighbor* (KNN), que também é um classificador, foi utilizado para estimar a pontuação do atleta com base nas distâncias entre o dado de entrada e os dados rotulados. O *PCA* também foi utilizado em Leo, D'Orazio e Trivedi (2009) e Sato *et al.* (2015) para reduzir a dimensionalidade dos dados.

Em Sato *et al.* (2015), a trajetória densa é utilizada como parâmetro de entrada do algoritmo *k-means*. O algoritmo agrupa as características das trajetórias que são usadas para criação de um histograma. Zecha e Lienhart (2015) também utilizam o *k-means* para agrupar as características das articulações.

Para avaliar o sincronismo de dois atletas durante a realização de saltos ornamentais, Ding *et al.* (2008) utilizam o algoritmo *RankBoost* (FREUND *et al.*, 2003) para construir a função de avaliação. Segundo os autores, a pontuação absoluta não pode ser utilizada devido às diferentes escalas de avaliação e variação dos juízes. Para contornar o problema, a pontuação relativa é usada, já que esta não sofre influência das diferentes escalas, fazendo com que se trate de um problema de ranqueamento.

3.4.3 Redes neurais convolucionais

Essa subseção apresenta os artigos que utilizaram redes neurais convolucionais para rastrear pontos de interesse, como articulações, ou reconstruir o esqueleto dos atletas. Também são apresentados *frameworks* com redes previamente treinadas. Essa técnica foi utilizada em sete dos artigos revisados (21,8%).

Para estimar a quantidade de passos dados por um atleta durante uma corrida de cem metros, Yagi *et al.* (2018) utilizaram o *OpenPose*[7] para reconstruir o esqueleto do corredor. Esse *framework* foi proposto em Cao *et al.* (2018) e permite identificar as articulações de múltiplas pessoas em uma mesma cena sem a necessidade de pré-processamento.

Em Cronin *et al.* (2019), os autores treinaram uma rede utilizando a biblioteca *DeepLabCut*[8] para identificar as articulações dos atletas durante a prática de corrida subaquática. Esse *framework* foi proposto com o objetivo de estimar a pose de animais, podendo ser treinado com um pequeno conjunto de imagens (MATHIS *et al.*, 2018). Nesse mesmo artigo, foram utilizados 500 quadros para o treinamento da rede.

Zecha *et al.* (2018) utilizaram a biblioteca *MobileNet*[9] (SANDLER *et al.*, 2018) com o objetivo de detectar a posição de esquiadores. Após essa etapa, os autores utilizaram uma *Convolutional Pose Machines*[10](WEI *et al.*, 2016) para estimar a pose dos atletas. Já para mensurar a força realizada em diferentes momentos antes do salto, os autores testaram diferentes configurações de uma *Temporal Convolutional Network* [11] (TCN) (BAI; KOLTER; KOLTUN, 2018), uma rede utilizada para a análise de sequências.

Com objetivo de medir a taxa de braçadas durante a natação, Victor *et al.* (2017) utilizaram uma rede neural convolucional para identificar o movimento dos atletas. Porém, ao invés de detectar o momento exato em que um evento ocorre (quadro em que o nadador submerge o braço na água e completa o movimento), os autores utilizaram as informações dos quadros anteriores e posteriores para estimar uma janela de confiança.

Em Parmar e Morris (2017), o objetivo do estudo foi atribuir nota à movimentos de esportes como ginástica artística, salto ornamental e patinação artística. Para tanto, os autores utilizaram uma *3D Neural Network* (C3D) (TRAN *et al.*, 2015) para extrair

7 ⟨https://github.com/CMU-Perceptual-Computing-Lab/openpose⟩
8 ⟨http://www.mousemotorlab.org/deeplabcut⟩
9 ⟨https://github.com/tensorflow/models/tree/master/research/slim/nets/mobilenet⟩
10 ⟨https://github.com/shihenw/convolutional-pose-machines-release⟩
11 ⟨https://github.com/locuslab/TCN⟩

características dos vídeos. Após essa etapa, três algoritmos são testados para atribuir a nota aos movimentos: SVR, *Long Short-Term Memory* (LSTM), e a combinação de SVR com LSTM.

No artigo de Yu *et al.* (2019), os autores utilizaram uma rede neural convolucional proposta em Zhou *et al.* (2017) para estimar o esqueleto tridimensional de uma pessoa utilizando como base uma imagem com somente duas dimensões. Após realizar uma análise cinemática dos movimentos, os dados são utilizados para estimar a fadiga de operários da construção civil.

Xiang *et al.* (2018) propõem um novo algoritmo para pontuar a execução de saltos ornamentais. A primeira parte do algoritmo consiste em dividir as etapas do salto utilizando uma *Encoder-Decoder Temporal Convolutional Network* (ED-TCN) (LEA *et al.*, 2017). Após essa etapa, quatro redes *Pseudo 3D* (P3D) (QIU; YAO; MEI, 2017) são treinadas para extrair as características. Finalmente, um algoritmo de regressão como o SVR é utilizado para atribuir a nota.

Wang *et al.* (2019) utilizaram diversas abordagens para avaliar a qualidade dos movimentos realizados por atletas de esqui estilo livre:

Para identificar a posição do esquiador no primeiro quadro do vídeo, os autores utilizaram uma variação da rede R-CNN (*Regions with Convolutional Neural Network*) (GIRSHICK *et al.*, 2014; REN *et al.*, 2017). Esse algoritmo é composto de duas etapas: na primeira, a rede procura por possíveis regiões de interesse; já a segunda etapa é responsável por classificar se há ou não humanos em tais regiões (WANG *et al.*, 2019).

A próxima etapa foi utilizar uma *Siamese Region Proposal Network* (Siamese-RPN)[12] (LI *et al.*, 2018; ZHANG; PENG, 2019) para rastrear o atleta nos quadros subsequentes. Segundo os autores, essa arquitetura utiliza uma imagem de menor resolução espacial contendo o elemento a ser rastreado e uma imagem completa da cena contendo a região de interesse, no caso, os quadros subsequentes do vídeo. As duas imagens são processadas e o resultado da rede é a posição do elemento buscado (WANG *et al.*, 2019).

Caso a qualidade do rastreamento seja menor que um limiar, um sistema de verificação é ativado. A rede utilizada nessa tarefa é uma modificação de uma *Multi-domain Convolutional Neural Network* (MDNet) (JUNG *et al.*, 2018). Caso a verificação não encontre a região de interesse, o algoritmo de detecção é executado novamente para toda a imagem (WANG *et al.*, 2019).

[12] ⟨https://github.com/researchmm/SiamDW⟩

Como última etapa, Wang *et al.* (2019) apresentam um novo algoritmo para extrair características da pose do atleta durante o movimento e refinar a posição de pontos chaves. O algoritmo utiliza a saída de uma rede convolucional, capaz de detectar os pontos de interesse (partes do corpo) e realiza uma interpolação bilinear para melhorar a precisão do rastreamento. Diferentemente de Xiao, Wu e Wei (2018), os autores utilizam uma sequência maior de quadros para suavizar a trajetória.

3.5 Amostras e conjuntos de dados

Foram extraídos dos artigos informações sobre a resolução espacial, frequência de captura e quantidade de câmeras utilizadas. Esta seção apresenta os resultados obtidos.

A resolução mais utilizada nos estudos foi de 1920 x 1080 px (*Full HD*) (4 artigos, 15%) (ZECHA *et al.*, 2018; GASTEL *et al.*, 2014; EVANS *et al.*, 2018; LEO; D'Orazio; TRIVEDI, 2009), seguida pela resolução de 1280 x 720 px (*HD*) (3 artigos, 9,3%) (EL-SALLAM *et al.*, 2013; MUKAI; ASANO; HARA, 2011; BARONE *et al.*, 2016). A maior resolução utilizada foi de 1920 x 1080 px (ONG; HARRIS; HAMILL, 2017), já a menor foi de 580 x 480 px (CRONIN *et al.*, 2019). Victor *et al.* (2017) não especificam o tamanho da imagem original, apenas informam que os fragmentos das imagens processadas possuem o tamanho de 128 x 48 px.

A frequência de captura mais comum foi de 30 quadros por segundo, os vídeos utilizados por cinco artigos (15,6%) possuem tal frequência (GASTEL *et al.*, 2014; GADE; LARSEN; MOESLUND, 2017; BARONE *et al.*, 2016; KOPOREC *et al.*, 2018; YU *et al.*, 2019). A taxa de captura de 50Hz foi a segunda mais utilizada (4 artigos, 12,5%) (ZECHA *et al.*, 2018; ZECHA; LIENHART, 2015; VICTOR *et al.*, 2017; EL-SALLAM *et al.*, 2013). A maior taxa de captura encontrada foi a de 240Hz, utilizada por Nagano *et al.* (2017). Já a taxa de 25Hz, a menor encontrada, foi utilizada por Leo, D'Orazio e Trivedi (2009), Ong, Harris e Hamill (2017) e Ceseracciu *et al.* (2011).

Quatorze dos artigos encontrados (43,7%) utilizam as imagens de apenas uma câmera para avaliar o desempenho dos atletas (CHU; SITUMEANG, 2017; YAGI *et al.*, 2018; JINYAN; GUANLEI; YU, 2013; GADE; LARSEN; MOESLUND, 2017; SHA *et al.*, 2014; ZECHA; LIENHART, 2015; VICTOR *et al.*, 2017; BARONE *et al.*, 2016; KOPOREC *et al.*, 2018; CRONIN *et al.*, 2019). Os estudos que mais utilizaram câmeras

foram El-sallam *et al.* (2013) e Sheets *et al.* (2011). Nesses estudos, oito câmeras foram utilizadas para reconstruir um modelo tridimensional do atleta. A tabela 8 lista todos os artigos e os respectivos dados.

Tabela 8 – Resolução das imagens utilizadas (em *pixels*), frequência de captura das câmeras (em *hertz*) e número de câmeras utilizadas (*Número de câmeras inferido a partir do artigo).

Artigo	Resolução	Frequência	Câmeras
Chu e Situmeang (2017)	854x480	–	1
Yagi *et al.* (2018)	640x320	–	1
Zecha *et al.* (2018)	1920x1080	50Hz	2*
Gastel *et al.* (2014)	1920x1080	30Hz	2
Jinyan, Guanlei e Yu (2013)	–	–	1
Gade, Larsen e Moeslund (2017)	640x480	30Hz	1
Evans *et al.* (2018)	1920x1080	180Hz	5
Leo, D'Orazio e Trivedi (2009)	1920x1080	25Hz	6
Sato *et al.* (2015)	–	–	1*
Sha *et al.* (2014)	–	–	1
Zecha e Lienhart (2015)	720x576	50Hz	1
Victor *et al.* (2017)	128x48	50Hz	1
El-sallam *et al.* (2013)	1280x720	50Hz	8
Ding *et al.* (2008)	–	–	1*
Parmar e Morris (2017)	–	–	1*
Mukai, Asano e Hara (2011)	1280x720	–	2
Barone *et al.* (2016)	1280x720	30Hz	1
Ong, Harris e Hamill (2017)	1384x1036	25Hz	2
Mehrizi *et al.* (2018)	720x480	–	2
Koporec *et al.* (2018)	640x480	30Hz	1
Sheets *et al.* (2011)	640x640	200Hz	8
Cronin *et al.* (2019)	580x480	60Hz	1
Díaz-Pereira *et al.* (2014)	–	–	1*
Ceseracciu *et al.* (2011)	720x576	25Hz	6
Yu *et al.* (2019)	640x480	30Hz	2
Mehrizi *et al.* (2017)	720x480	–	2
Wang *et al.* (2019)	–	–	1*
Wu *et al.* (2019)	–	–	6
Shin e Ozawa (2008)	–	–	1*
Boonim e Sanguansat (2018)	–	–	1*
Xiang *et al.* (2018)	–	–	1*
Nagano *et al.* (2017)	–	240Hz	1*

Fonte – Daniel Freire Tsuha, 2019

As amostras utilizadas nos estudos foram variadas. Alguns artigos apresentavam em detalhes o número de participantes e o protocolo de captura dos dados. Já outros, utilizaram conjuntos de dados públicos ou gravações de jogos ou partidas. Shin e Ozawa (2008) utilizaram uma amostra de 400 estudantes em seus experimentos, sendo a maior

amostra encontrada. A tabela 9 mostra em detalhes o número de participantes ou a quantidade de vídeos utilizados em cada estudo.

Tabela 9 – Amostras utilizadas nos estudos para verificar a precisão da técnica ou treinar o algoritmo de reconhecimento.

Artigo	Amostra
Chu e Situmeang (2017)	6 vídeos, 558 jogadas
Yagi *et al.* (2018)	29 atletas
Zecha *et al.* (2018)	225 vídeos
Gastel *et al.* (2014)	7 atletas, 3 execuções em velocidades diferentes
Jinyan, Guanlei e Yu (2013)	–
Gade, Larsen e Moeslund (2017)	1 atleta, 2 exercícios diferentes
Evans *et al.* (2018)	18 atletas, 10 execuções
Leo, D'Orazio e Trivedi (2009)	4 atletas analisados por 2 câmeras
Sato *et al.* (2015)	30 atletas, 5 execuções
Sha *et al.* (2014)	5 atletas, 8 execuções
Zecha e Lienhart (2015)	14 atletas, 30 vídeos
Victor *et al.* (2017)	40 atletas
El-sallam *et al.* (2013)	5 atletas, vários exercícios
Ding *et al.* (2008)	30 vídeos
Parmar e Morris (2017)	Vários conjuntos de dados de diferentes esportes
Mukai, Asano e Hara (2011)	4 atletas
Barone *et al.* (2016)	6 atletas, 3 execuções em velocidades diferentes
Ong, Harris e Hamill (2017)	10 atletas
Mehrizi *et al.* (2018)	12 atletas, várias tentativas
Koporec *et al.* (2018)	12 atletas
Sheets *et al.* (2011)	7 atletas
Cronin *et al.* (2019)	21 atletas
Díaz-Pereira *et al.* (2014)	8 atletas, 10 movimentos, de 5 a 7 execuções
Ceseracciu *et al.* (2011)	5 atletas
Yu *et al.* (2019)	–
Mehrizi *et al.* (2017)	12 atletas, várias tentativas
Wang *et al.* (2019)	30 atletas, 63 vídeos
Wu *et al.* (2019)	–
Shin e Ozawa (2008)	400 estudantes, 5 etapas, 240 vídeos
Boonim e Sanguansat (2018)	–
Xiang *et al.* (2018)	370 vídeos de 4 segundos
Nagano *et al.* (2017)	8 horas de vídeo, 2 tipos de exercícios

Fonte – Daniel Freire Tsuha, 2019

3.6 Formas de avaliação

Uma das formas de verificar o desempenho dos algoritmos é comparar os resultados obtidos com outras formas de avaliação e/ou equipamentos. Dos artigos analisados, 13

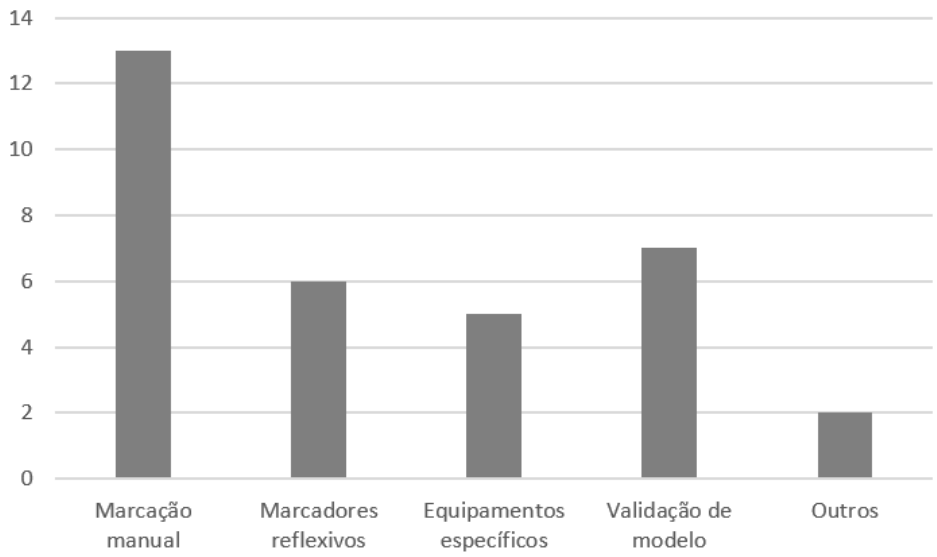
(40,6%) compararam os resultados da técnica desenvolvida com a marcação manual realizada por um ou mais operadores (YAGI *et al.*, 2018; JINYAN; GUANLEI; YU, 2013; LEO; D'Orazio; TRIVEDI, 2009; SHA *et al.*, 2014; VICTOR *et al.*, 2017; MUKAI; ASANO; HARA, 2011; CRONIN *et al.*, 2019; DÍAZ-PEREIRA *et al.*, 2014; CESERACCIU *et al.*, 2011; WU *et al.*, 2019; SHIN; OZAWA, 2008; BOONIM; SANGUANSAT, 2018; XIANG *et al.*, 2018).

Já a comparação com sistemas de marcadores reflexivos foi usada por Evans *et al.* (2018), El-sallam *et al.* (2013), Ong, Harris e Hamill (2017), Mehrizi *et al.* (2018), Nagano *et al.* (2017) e Mehrizi *et al.* (2017) (6 artigos, 18,7%). Essa técnica é considerada o padrão-ouro para a reconstrução de movimento. Porém, devido às limitações do equipamento, essa abordagem não pode ser utilizada em todos os contextos, o que explica o menor uso do método.

Dependendo da atividade estudada, pode-se utilizar sensores específicos para capturar os dados, cinco estudos usam essa abordagem (15,6%). Em Yagi *et al.* (2018), os autores utilizaram um equipamento de medição de velocidade baseado em *lasers*; já em Gade, Larsen e Moeslund (2017), é usado um sistema móvel de teste de exercício cardiopulmonar para saber quanto oxigênio foi consumido pelo atleta; Evans *et al.* (2018) utilizaram plataformas de pressão para calcular o tempo e a frequência de passos durante a corrida; Zecha *et al.* (2018) também utilizaram plataformas de pressão para estimar a força realizada pelo esquiador antes do salto; por fim, Yu *et al.* (2019) fizeram o uso de um acelerômetro para estimar a fadiga de operários.

Outras abordagens utilizadas como referência para avaliar os desempenhos das ferramentas foram: comparação entre os resultados obtidos entre uma e várias câmeras, técnica utilizada por Barone *et al.* (2016); e a análise gráfica utilizada por Sato *et al.* (2015), em que, após a extração de características dos vídeos, os autores analisaram graficamente o comportamento das variáveis de atletas amadores e experientes. Por fim, sete artigos (21,8%) avaliaram os modelos treinados para a realização de tarefas específicas (CHU; SITUMEANG, 2017; ZECHA; LIENHART, 2015; DING *et al.*, 2008; PARMAR; MORRIS, 2017; KOPOREC *et al.*, 2018; SHEETS *et al.*, 2011; GASTEL *et al.*, 2014). A figura 20 mostra a frequência das formas de avaliação.

Figura 20 – Formas de avaliação.



Fonte – Daniel Freire Tsuha, 2020

3.7 Limitações

Entre as limitações e os próximos passos apresentados, a melhora da precisão do algoritmo foi a mais citada (10 artigos, 31,5%) (CHU; SITUMEANG, 2017; YAGI *et al.*, 2018; ZECHA *et al.*, 2018; GADE; LARSEN; MOESLUND, 2017; LEO; D'Orazio; TRIVEDI, 2009; SHA *et al.*, 2014; EL-SALLAM *et al.*, 2013; KOPOREC *et al.*, 2018; CRONIN *et al.*, 2019; YU *et al.*, 2019). Zecha *et al.* (2018) e El-sallam *et al.* (2013) enumeram especificamente a remoção de ruído para melhorar o desempenho do algoritmo.

Gade, Larsen e Moeslund (2017), Sha *et al.* (2014), Barone *et al.* (2016), Koporec *et al.* (2018), Cronin *et al.* (2019), Ceseracciu *et al.* (2011) e Yu *et al.* (2019) apresentam como próximo passo da pesquisa a extração de mais características dos vídeos para a análise (7 artigos, 21,8%). A extração de mais métricas em trabalhos futuros é apontado por Zecha *et al.* (2018), Leo, D'Orazio e Trivedi (2009), Victor *et al.* (2017), Mukai, Asano e Hara (2011), Barone *et al.* (2016) e Yu *et al.* (2019) (6 artigos, 18,7%).

A ampliação para outros esportes ou situações foram listados como próximos passos por seis artigos (18,7%) (JINYAN; GUANLEI; YU, 2013; ZECHA; LIENHART, 2015; MEHRIZI *et al.*, 2018; KOPOREC *et al.*, 2018; CESERACCIU *et al.*, 2011; MEHRIZI *et al.*, 2017). Já a melhoria do algoritmo para a utilização em ambientes e situações de menor controle foi apresentada como objetivo futuro por Gade, Larsen e Moeslund (2017), Evans

et al. (2018), Mehrizi *et al.* (2018), Koporec *et al.* (2018), Shin e Ozawa (2008) e Mehrizi *et al.* (2017) (6 artigos, 18,7%).

Por fim, Díaz-Pereira *et al.* (2014) apresentam como próximos passos da pesquisa a automatização de algumas etapas do processo e a utilização de mais câmeras para analisar a movimentação dos atletas.

3.8 Discussão

Após a análise dos artigos, é possível extrair algumas conclusões sobre a avaliação esportiva baseada em vídeo. A primeira delas é que, dada as especificações das câmeras utilizadas, é possível afirmar que as imagens gravadas por um *smartphone* modelo de entrada podem ser usadas para analisar o desempenho de atletas, já que a maioria desses modelos apresentam uma resolução mínima de 1920 x 1080 px e realizam capturas de 30 quadros por segundo. Para os esportes e métricas apresentados, tais requisitos são suficiente para obter medições confiáveis para a maioria dos casos. Aparelhos de celular mais sofisticados permitem a gravação de vídeos com uma maior taxa de quadros por segundo, o que pode aumentar a precisão de algumas análises. Isso implica no barateamento do processo e diminui a necessidade de aquisição de equipamentos específicos. Entretanto, algumas métricas, como tempo de reação de um atleta, podem demandar equipamentos com maiores resoluções e frequências de captura.

Com exceção da reconstrução do modelo tridimensional do atleta, não foi possível encontrar uma metodologia em comum entre os artigos analisados. A diversidade de ambientes, movimentos e métricas demandam soluções específicas para cada contexto. As formas de avaliação também foram variadas, as ferramentas consideradas como padrão-ouro são diferentes para cada aplicação. A avaliação estatística dos resultados também depende da métrica extraída e da metodologia utilizada.

Para tentar contornar esse problema, alguns estudos utilizaram redes neurais convolucionais já treinadas para esqueletizar os atletas e encontrar automaticamente a posição das principais articulações do corpo. Ferramentas como *PoseNet*[13], *OpenPose*[14] e

[13] ⟨https://github.com/tensorflow/tfjs-models/tree/master/posenet⟩
[14] ⟨https://github.com/CMU-Perceptual-Computing-Lab/openpose⟩

DensePose[15] podem ser utilizadas nesses contextos, diminuindo a necessidade de processos específicos para diferentes esportes.

Uma das razões pelas quais corrida e caminhada terem sido os esportes mais estudados pode ser atribuído a simplicidade necessária para sua prática. Percebe-se que há predominância do uso de técnicas de segmentação devido à natureza dos dados (vídeos de práticas esportivas). Mapeamos o uso de técnicas mais avançadas de processamento gráfico. Embora o uso de redes neurais convolucionais apareça direta ou indiretamente em pelo menos um quinto dos estudos, a quantidade de dados necessários para o bom funcionamento dessas redes é muito grande, o que geralmente é problemático, uma vez que as amostras e atletas utilizados nos vídeos são limitados em número.

Se as redes neurais convolucionais se mostrarem precisas o suficiente para reconstruir o movimento de atletas, então existe a possibilidade para o desenvolvimento de um *framework* que possa abranger diversos esportes. A partir do rastreamento das articulações, diversas métricas físicas e técnicas podem ser extraídas, sendo esse, um possível próximo passo no desenvolvimento de ferramentas para avaliação física.

Quanto aos conjuntos de dados, não há padronização. Parece haver uma adaptação das necessidades dos pesquisadores de acordo com o número de atletas e o equipamento disponível. No entanto, observa-se que na maior parte dos trabalhos não foram utilizados equipamentos especiais, como uma câmera de grande resolução ou que fizesse a captura em uma frequência muito alta.

Observou-se que a maioria dos estudos encontrados utilizou a marcação manual como forma de avaliação. Sendo notável o espaço para o desenvolvimento de modelos que permitam uma validação mais automática das técnicas, não apenas para facilitar a análise, mas também para aumentar a confiabilidade dos dados.

3.9 Conclusão

Realizou-se uma revisão sistemática da literatura usando os protocolos estabelecidos no trabalho de Kitchenham (2004) sobre o tema "analisando o desempenho dos atletas a partir de imagens capturadas por câmeras de vídeo". Os artigos publicados a partir de 2008 foram considerados. Dos 5461 documentos que apareceram na pesquisa inicial, apenas foram selecionados 32 nesta revisão.

[15] ⟨http://densepose.org/⟩

Com o tempo, observou-se um aumento do interesse da comunidade científica pelo assunto, o que pode ser evidenciado pelo crescente número de artigos por ano catalogados neste estudo.

Entre os esportes analisados, corrida/caminhada foi a mais estudada, aparecendo em 10 dos 32 artigos. Entre as técnicas de visão computacional adotadas, a subtração de segundo plano parece ser a mais popular entre os pesquisadores, aparecendo em 17 estudos. Quanto aos conjuntos de dados, embora exista muita variação, a resolução *full HD* (1920 x 1080 px) foi a mais utilizada, aparecendo em 15% dos artigos, e a frequência de captura mais comum foi de 30 quadros por segundo, finalmente, 43,7% (14 artigos) usaram apenas uma câmera para capturar imagens. Quanto às formas de avaliação, a comparação com a marcação manual foi o método mais comum, aparecendo em 13 estudos. Dos artigos que citaram seus próximos passos, a maior parte (10 estudos) mencionou melhora na precisão dos algoritmos.

Com exceção da técnica de reconstrução tridimensional, não foi possível encontrar um processo padronizado de captura, processamento e análise. Como cada esporte apresenta suas peculiaridades (tipos de movimentos, métricas e ambientes), os estudos desenvolveram metodologias específicas para cada contexto.

Dado a abrangência do assunto, ainda existem diversas lacunas a serem solucionadas e esportes a serem estudados, sendo esse um campo de estudo com diversas oportunidades. A utilização de redes neurais convolucionais tem se mostrado uma ferramenta promissora para a análise de movimento e sua utilização deve ser ampliada.

Em resumo, esta é uma área de pesquisa interdisciplinar, com muito espaço para avanços, tanto no campo das técnicas computacionais quanto no campo dos atletas e treinadores. Ainda existem muitos problemas a serem resolvidos, fornecendo um vasto campo de pesquisa que tende a evoluir ao longo dos anos. Enfatiza-se que o maior problema detectado nesta revisão foi que, quanto mais automatizado o método, mais difícil se torna manter a precisão necessária para uma análise eficiente. Portanto, encontrar um equilíbrio entre esses dois requisitos é o ponto principal de qualquer técnica de análise automática de desempenho para atletas.

4 Metodologia

O presente capítulo descreve o processo para estimar a posição de escaladores durante provas de escalada de velocidade. A seção 4.1 apresenta o conjunto de dados utilizados e o processo de rotulação das imagens e as seções 4.2 e 4.3 o processo de detecção das agarras e dos escaladores respectivamente. Já a seção 4.4 apresenta os passos para mapear as coordenadas da imagem para as posições no mundo real e, por fim, a seção 4.5 apresenta como o desempenho do sistema foi mensurado.

4.1 Conjunto de dados

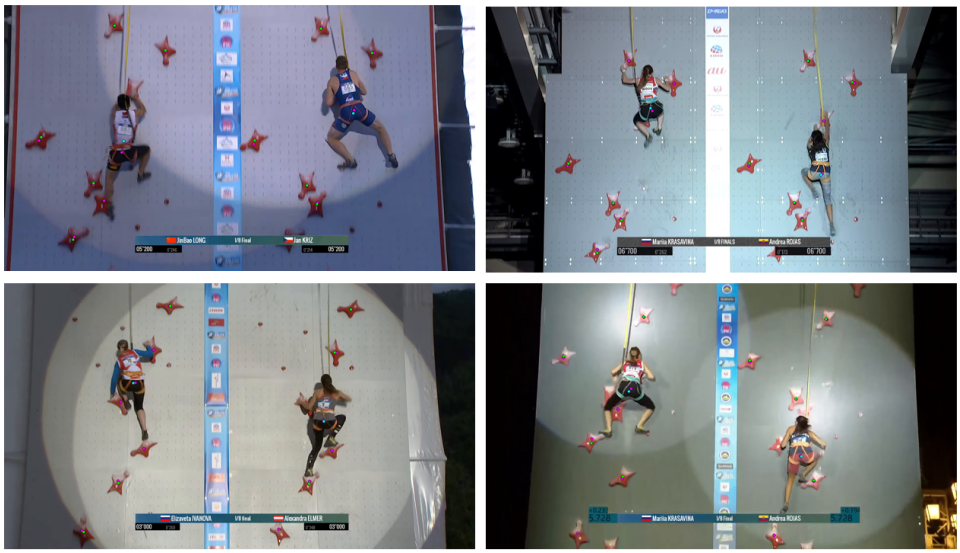
A primeira etapa da criação do conjunto de dados foi a escolha de gravações de campeonatos de escalada de velocidade disponíveis na internet. Dentre os diversos vídeos disponíveis, quatro foram selecionados para serem utilizados no presente projeto, os vídeos foram escolhidos de forma que as condições de iluminação e posicionamento da câmera fossem o mais diversas o possível. A próxima etapa foi cortar os vídeos de forma a manter somente as corridas e remover todo o conteúdo irrelevante como comentários, preparação dos atletas e *replays*. Devido às particularidades de cada competição, cada gravação possui diferentes quantidades de corridas e, ao final do processo, 80 corridas foram extraídas.

A resolução das imagens utilizadas foi de 1920 x 1080 *pixels* e a frequência de captura das câmeras era de 30 *frames* por segundo. O conjunto de dados possui 20.066 *frames*, totalizando aproximadamente 668 segundos de vídeo. Os vídeos foram baixados utilizando a ferramenta *youtube-dl*[1]. Entretanto, devido às múltiplas câmeras utilizadas, em alguns vídeos não é possível recuperar os primeiros momentos da corrida. A figura 21 mostra exemplos de *frames* das gravações utilizadas e os vídeos podem ser encontrados na integra na internet[2].

[1] ⟨https://github.com/ytdl-org/youtube-dl⟩

[2] ⟨https://youtu.be/i0vHEP6T7fw⟩
⟨https://youtu.be/y9ZQj3758mw⟩
⟨https://youtu.be/pJKVOWApsEU⟩
⟨https://youtu.be/BIcThugvCPQ⟩

Figura 21 – Exemplos de *frames* presentes no conjunto de dados.



Fonte – International Federation of Sport Climbing

4.2 Detecção das agarras

Após a seleção e recorte dos vídeos, realizou-se a detecção das agarras. Nessa etapa foi utilizada a rede neural convolucional *YOLO* descrita na subseção 2.5.1 e o processo de treinamento está descrito na subseção 4.2.2. Já o processo de construção do conjunto de dados de agarras está descrito na subseção 4.2.1.

4.2.1 Conjunto de dados de agarras

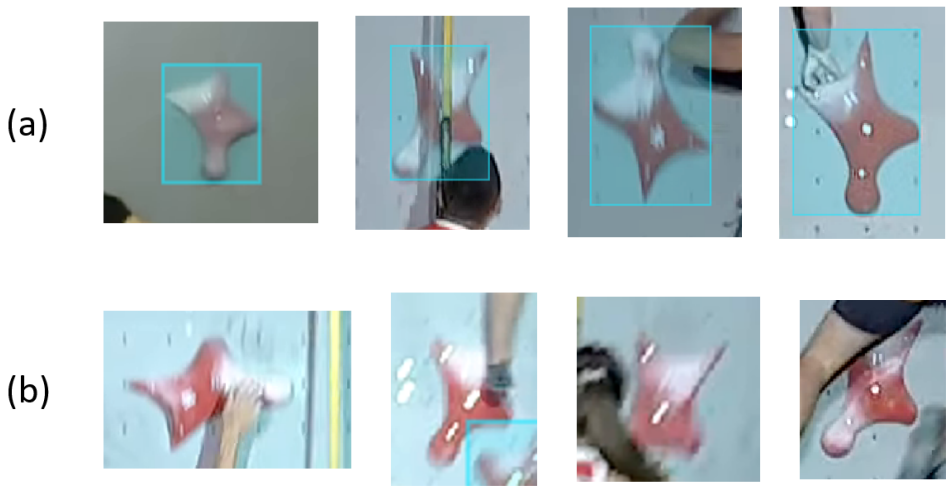
Para a criação do conjunto de dados de agarras 250 *frames* foram escolhidos, aleatoriamente, de cada um dos quatro vídeos totalizando um total de mil *frames*. Após essa seleção, as agarras foram rotuladas manualmente, utilizando a ferramenta LabelBox[3], de modo que, ao final do processo, o conjunto de dados contava com 7.165 exemplos de regiões retangulares contendo agarras.

Para que uma agarra fosse considerada válida, ela deveria estar completamente visível, qualquer agarra parcialmente ocluída ou cortada, foi ignorada, exceto quando a oclusão era causada pela fita de segurança presa ao atleta. Como o centro da região detectada foi utilizado para estimar a posição de uma agarra, o reconhecimento parcial deslocaria o centro da região de interesse, o que acarretaria no aumento do erro da estimativa

[3] Disponível em ⟨https://labelbox.com/⟩

final. A figura 22 mostra alguns exemplos de agarras consideradas e não consideradas durante o processo de rotulação.

Figura 22 – Exemplos de agarras: a) agarras selecionadas para treinamento, as agarras ocluídas parcialmente pela fita de segurança foram consideradas; b) exemplo de agarras ocluídas não consideradas para o treinamento, já que o centro da região retangular destoaria ainda mais da região utilizada como referência para os cálculos.



Fonte – International Federation of Sport Climbing

4.2.2 Treinamento da rede YOLO

Foi utilizada a quarta versão da rede *YOLO* implementada no *framework* Darknet[4]. Para reduzir a quantidade de dados necessários e o tempo de treinamento, usou-se um conjunto de pesos de uma rede pré-treinada para detectar diferentes classes de objetos. A vantagem dessa abordagem, conhecida como transferência de conhecimento, é o reaproveitamento de filtros de extração de características genéricos que foram aprendidos a partir de outros objetos (SHAO; ZHU; LI, 2015; KASHIPAREKH *et al.*, 2019).

Para o treinamento da rede também foram utilizadas técnicas de aumento de dados, que consistem em gerar variações das imagens do conjunto de dados para melhorar a capacidade de generalização da rede (GAO; CAI; MING, 2020). Das opções disponíveis, as utilizadas foram: rotação das imagens, aplicação de efeitos de borramento (*blur*) e variação de brilho, saturação e cor. A rede foi treinada por 640 épocas a resolução de saída foi de 640 x 640 *pixels*. Para os demais parâmetros da rede, foram mantidos os valores padrão. O

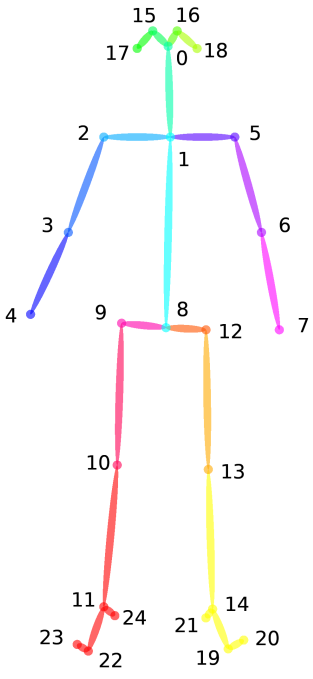
[4] Disponível em ⟨https://github.com/pjreddie/darknet⟩

treinamento da rede foi realizado com 800 imagens do conjunto de dados e 200 imagens foram utilizadas para a avaliação da rede.

4.3 Detecção dos escaladores

Para realizar a detecção dos escaladores foi utilizada a rede *OpenPose* (CAO *et al.*, 2019), detalhada na subseção 2.5.2. Para cada atleta detectado pela rede é retornado um conjunto de 25 pontos que representam o esqueleto do atleta. Como referência para estimar a posição dos escaladores, foi utilizado o centro do quadril, tal ponto é uma aproximação do centro de massa do corpo de uma pessoa e é comumente usado para análises esportivas. A figura 23 mostra a esqueletização gerada pela rede.

Figura 23 – Esqueletização gerada pelo *OpenPose*: o ponto 8 foi utilizado como estimativa do centro de massa do atleta.



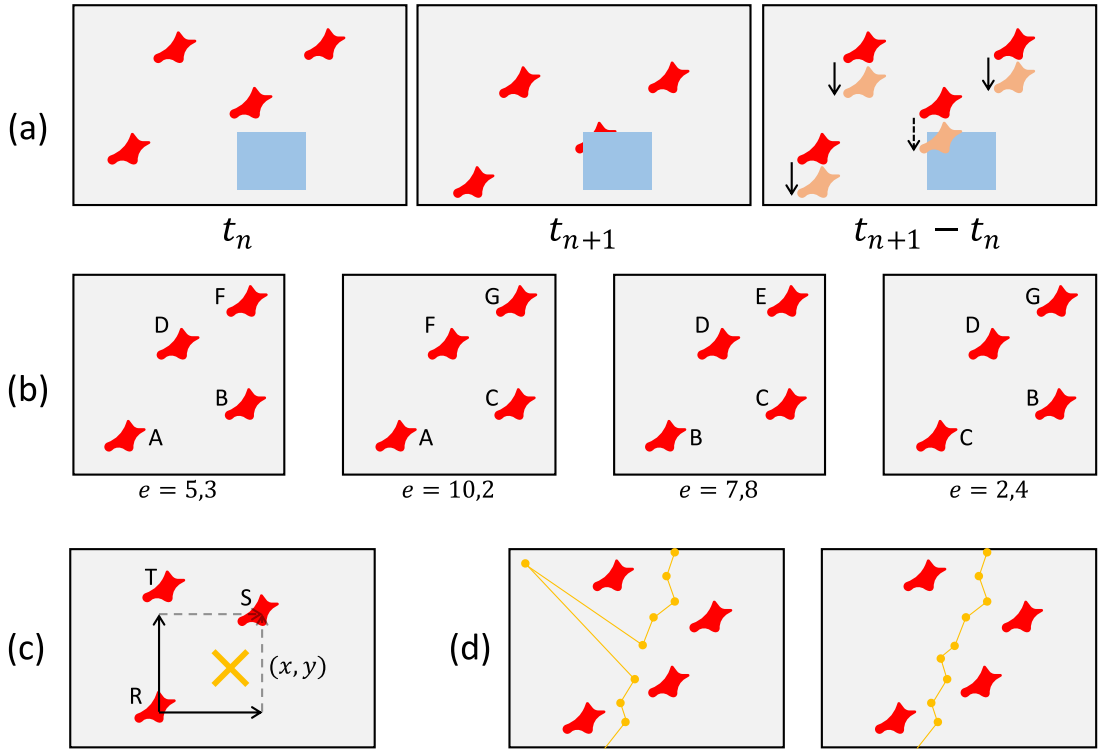
Fonte – Documentação do OpenPose, disponível em ⟨https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_02_output.html⟩.

4.4 Mapeamento dos dados

Após obter a posição das agarras e dos escaladores na imagem, o próximo passo consiste em estimar a posição do atleta no mundo real. Para alcançar tal objetivo, o

processo foi divido em quatro etapas: a estimativa de posição de agarras ocluídas, a identificação das agarras, a estimativa da posição do escalador no mundo real e o pós-processamento. Cada uma dessas etapas será descrita a seguir e a figura 24 apresenta de forma resumida cada uma delas.

Figura 24 – Processo de mapeamento das agarras: a) Propagação das agarras: comparando o deslocamento das agarras de dois *frames* consecutivos, é possível estimar a posição das agarras ocluídas; b) Estimativa de erro: para identificar as agarras, compara-se diversas combinações possíveis e calcula-se o erro e para cada combinação; c) Mapeamento das agarras: após a identificação das agarras, é realizada a estimativa da posição do centro de massa do atleta (X amarelo); d) Pós processamento: no pós processamento, os pontos que foram detectados erroneamente são corrigidos.



Fonte – Daniel Freire Tsuha, 2023

4.4.1 Estimativa de posição de agarras ocluídas

Em alguns momentos, a quantidade de agarras detectadas em uma imagem pode não ser suficiente para identificá-las corretamente. Porém, devido à natureza dinâmica dos vídeos, é possível utilizar os dados do *frame* anterior para estimar a posição de uma agarra no *frame* seguinte. Essa abordagem é especialmente útil nos casos em que o corpo do atleta oclui um considerável número de agarras.

A abordagem implementada utiliza-se do fato que a câmera se desloca verticalmente e as agarras são detectadas principalmente nas partes superiores e inferiores das imagens, já que normalmente os escaladores se encontram da região central. Sendo assim, é possível calcular a diferença no deslocamento entre duas imagens consecutivas e replicar a posição das agarras detectadas no *frame* anterior. Dessa forma, mesmo que o atleta esteja obstruindo uma agarra, ainda é possível estimar a sua posição. O algoritmo é dividido em duas etapas: a primeira consiste no cálculo do deslocamento vertical, enquanto na segunda parte é realizado a estimativa da posição das agarras.

O algoritmo 1 apresenta os passos executados para se obter o deslocamento entre dois *frames* consecutivos. As funções BordaSuperior e BordaInferior criam uma margem de altura d nas bordas superiores e inferiores da imagem respectivamente. Já a função DeslocamentoVertical desloca a imagem i *pixels* para baixo. Tal algoritmo deve ser executado para todos os quadros do vídeo. A figura 25 ilustra o processo descrito.

As bordas adicionadas nas partes superiores e inferiores da imagem são necessárias para eliminar a influência do movimento da câmera, já que essas regiões apresentam as maiores diferenças na comparação de dois *frames* consecutivos.

Algoritmo 1 Cálculo do deslocamento entre dois frames

d: tamanho da janela de busca
F_n: n-ésimo frame

$F_n \leftarrow \text{BordaSuperior}(F_n, d)$
$F_n \leftarrow \text{BordaInferior}(F_n, d)$

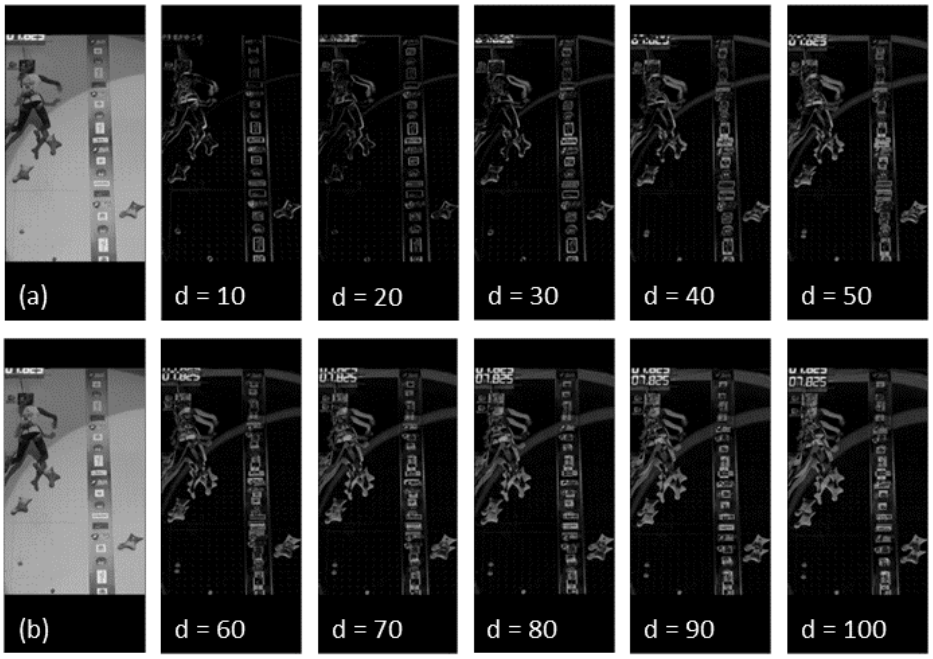
$menorSoma \leftarrow maxInt$
$deslocamento \leftarrow 0$
$i \leftarrow 0$
while $i < d$ **do**
    $F_{n-1} \leftarrow \text{BordaSuperior}(F_{n-1}, d)$
    $F_{n-1} \leftarrow \text{BordaInferior}(F_{n-1}, d)$
    $F_{n-1} \leftarrow \text{DeslocamentoVertical}(F_{n-1}, i)$

    $diferencas \leftarrow |F_n - F_{n-1}|$
    $soma \leftarrow \text{Soma}(diferencas)$

    if $soma < menorSoma$ **then**
        $menorSoma \leftarrow soma$
        $deslocamento \leftarrow i$
    $i \leftarrow i + 1$

return $deslocamento$

Figura 25 – Cálculo de deslocamento: (a) e (b) exemplificam dois *frames* consecutivos, as demais imagens representam a diferença absoluta entre os *frames* e o valor d representa o deslocamento vertical da imagem (b) em relação à imagem (a). Quanto mais *pixels* pretos, maior a sobreposição das imagens.



Fonte – International Federation of Sport Climbing

O processo para estimar a posição de uma agarra ocluída é descrito no algoritmo 2. A função DeslocaAgarra desloca as coordenadas de uma agarra d *pixels* para cima. Já a função AgarraMaisProxima é utilizada para saber qual é a agarra do *frame* anterior mais próxima de uma determinada agarra do *frame* atual. A figura 26 ilustra o processo.

O algoritmo foi executado para todos os pares de *frames* consecutivos. As agarras propagadas de um *frame* para o outro são consideradas como parâmetros de entrada da iteração seguinte, ou seja, mesmo que uma agarra seja ocluída por diversos quadros consecutivos, ainda sim é possível estimar a sua posição.

4.4.2 Identificação das agarras

Após o processamento inicial das agarras, o passo seguinte é identificá-las seguindo a notação definida na seção 2.1. A intuição do algoritmo apresentado é testar qual combinação de agarra melhor se encaixa com as agarras detectadas. Para tanto, primeiro estima-se a escala da imagem utilizando a posição das agarras detectadas e calcula-se uma medida de erro baseada nas dimensões reais da parede. O algoritmo 3 descreve os passos executados

Algoritmo 2 Estimativa da posição de agarras ocluídas

H_n: lista de agarras do frame n
h_n^m: m-ésima agarra do n-ésimo frame
delta: deslocamento entre os frames F_n e F_{n-1}

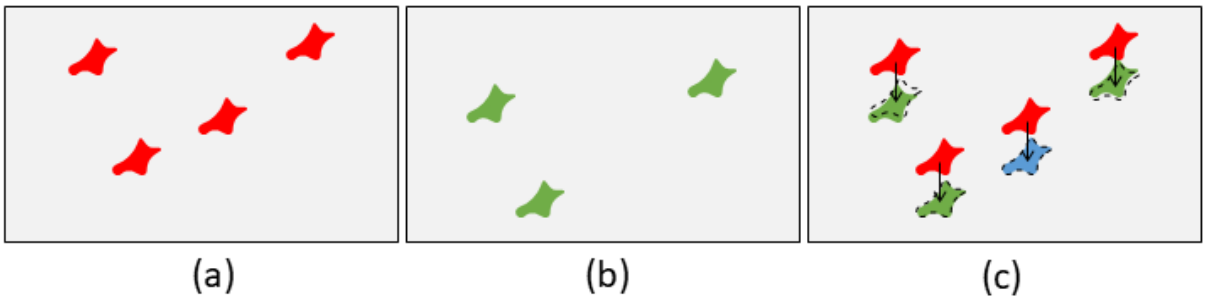
for $h_{n-1}^m \in H_{n-1}$ **do**
    $h_{n-1}^m \leftarrow$ DESLOCAAGARRA$(h_{n-1}^m, delta)$

$propagadas \leftarrow []$
for $h_n^m \in H_n$ **do**
    $h_{prox} \leftarrow$ AGARRAMAISPROXIMA(h_n^m, H_{n-1})
    $dist \leftarrow$ DISTANCIA(h_n^m, h_{prox})

    if $dist >$ altura da agarra h_n^m **then**
        $propagadas \leftarrow propagadas \cup [h_{prox}]$

return $H_n \cup propagadas$

Figura 26 – Estimativa da posição das agarras ocluídas: a) agarras detectadas no *frame* n_{t-1}; b) agarras detectadas no *frame* n_t; c) em tracejado a propagação das agarras de n_{t-1} em n_t, como a agarra em azul não possui um correspondente no *frame* atual, a posição será estimada a partir dos dados do *frame* anterior.

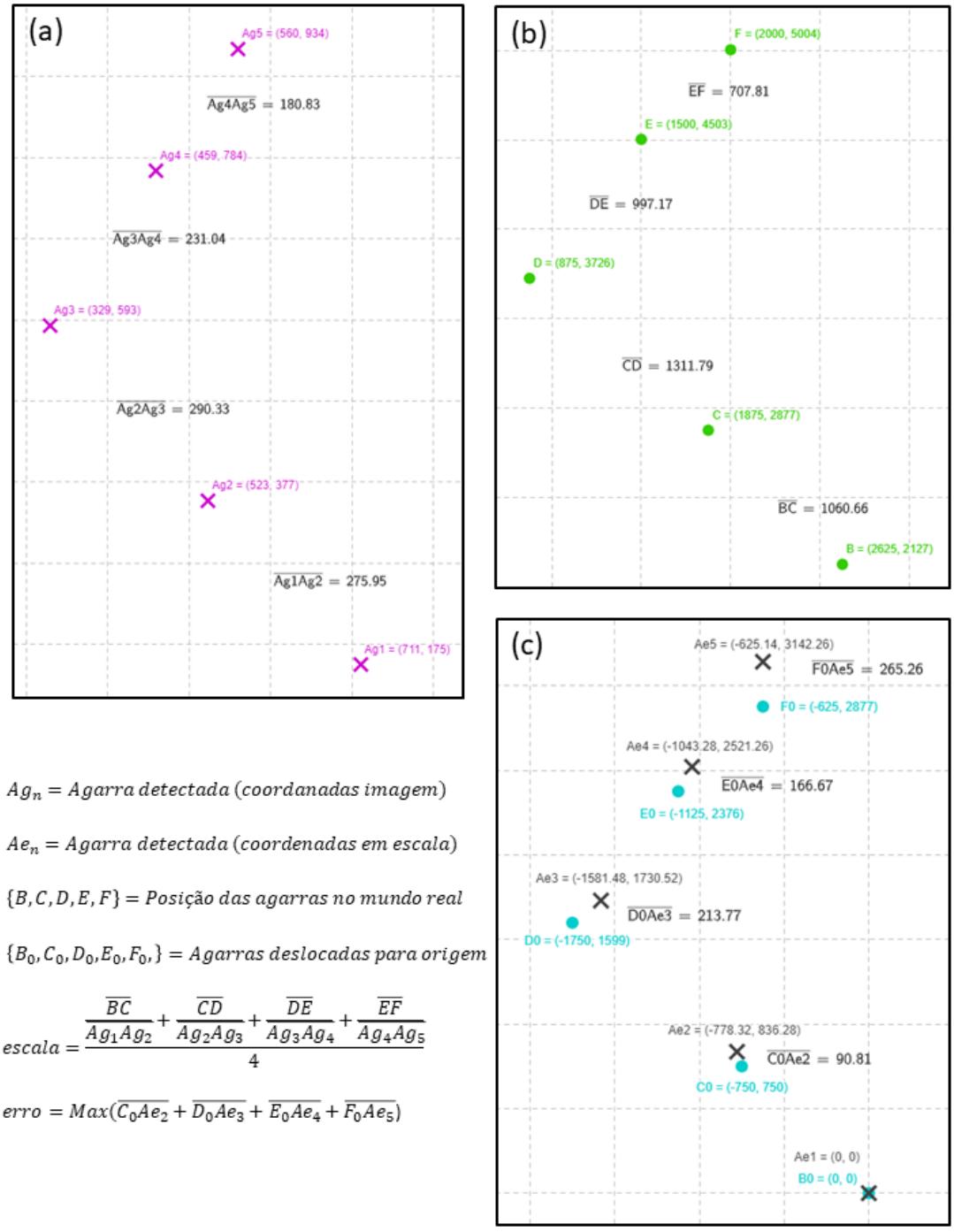


Fonte – Daniel Freire Tsuha, 2023

para identificar as agarras e a figura 27 ilustra o processo de forma visual. Os detalhes das funções são explicados a seguir.

O primeiro passo para identificar as agarras é criar todos os arranjos possíveis a serem testados. A função CANDIDATOS recebe como parâmetro uma lista contendo as informações de todas as agarras do mundo real e um parâmetro l que indica a quantidade de agarras detectadas. A função então cria todos os arranjos ordenados de l elementos de forma que a distância alfabética entre a primeira e a última letra seja menor que dez. Como só é possível ver uma fração da parede a cada imagem, não faz sentido testar combinações contendo agarras muito distantes, como por exemplo o arranjo $[A, B, T]$.

Figura 27 – Exemplo real do cálculo do erro de uma combinação de agarras: a) coordenadas das agarras detectadas e as distâncias entre as agarras consecutivas (px); b) coordenadas das agarras no mundo real e as distâncias entre as agarras consecutivas (mm), a escala é calculada a partir da média das distâncias; c) após aplicar a escala às agarras detectadas, desloca-se os dois conjuntos de pontos de forma que a agarra mais abaixo esteja na origem do plano, feito isso, calcula-se a distância entre as agarras detectadas e as agarras do mundo real, o erro é a maior distância entre a estimativa calculada e a posição da agarra no mundo real.



$Ag_n = Agarra\ detectada\ (coordanadas\ imagem)$

$Ae_n = Agarra\ detectada\ (coordenadas\ em\ escala)$

$\{B, C, D, E, F\} = Posição\ das\ agarras\ no\ mundo\ real$

$\{B_0, C_0, D_0, E_0, F_0,\} = Agarras\ deslocadas\ para\ origem$

$$escala = \frac{\frac{\overline{BC}}{\overline{Ag_1Ag_2}} + \frac{\overline{CD}}{\overline{Ag_2Ag_3}} + \frac{\overline{DE}}{\overline{Ag_3Ag_4}} + \frac{\overline{EF}}{\overline{Ag_4Ag_5}}}{4}$$

$$erro = Max(\overline{C_0Ae_2} + \overline{D_0Ae_3} + \overline{E_0Ae_4} + \overline{F_0Ae_5})$$

Fonte – Daniel Freire Tsuha, 2023

Algoritmo 3 Detectar a melhor combinação

H: lista de agarras detectadas
R: lista de todas as agarras em proporções do mundo real
l: número de agarras detectadas (cardinalidade de H)

$C \leftarrow$ CANDIDATOS(R, l)

$menorErro \leftarrow maxInt$
$melhorCombinacao \leftarrow []$
for $W_i \in C$ **do**
    $erro \leftarrow$ ERROCOMBINACAO(H, W_i)
    if $erro < menorErro$ **then**
        $menorErro \leftarrow erro$
        $melhorCombinacao \leftarrow w_i$
return $melhorCombinacao$

Após todas as combinações válidas serem geradas, o erro de cada uma delas é estimado e a combinação com o menor erro é escolhida para estimar a posição do atleta. A função ERROCOMBINACAO é explicada em detalhes no algoritmo 4. O erro de uma combinação consiste no maior valor medido entre a posição de uma agarra detectada e a respectiva agarra no mundo real. Ou seja o menor erro dentre todos os piores casos.

A escala da imagem é calculada a partir da média das proporções dos pares de agarras consecutivas. Para cada par de agarras detectadas em sequência, calcula-se o a distância entre elas, o mesmo processo é realizado para o par de agarras análogos em proporções reais, divide-se então distância real pela distância da imagem. Feito isso para todos os pares de agarra, é calculada a média desses valores. O pseudocódigo da função está descrito no algoritmo ESCALA.

Já a função DESLOCAPARAORIGEM tem por objetivo tornar os dados da imagem comparáveis com os dados do mundo real. A ideia do algoritmo é colocar as coordenadas das agarras detectadas na mesma escala das agarras do mundo real. Feito isso, move-se ambos os conjuntos de agarras para um espaço onde possam ser comparadas. Tal processo é feito de modo que as coordenadas da agarra mais ao sul seja a origem dos eixos, ou seja $(0, 0)$.

Uma vez que as agarras detectadas e as agarras do mundo real estão na mesma escala em no mesmo espaço, a função ERRO pode então calcular o erro de uma função. Para tanto mede-se a distância entre as agarras detectadas e as agarras análogas no mundo real. O maior valor é utilizado como erro de estimativa de uma combinação.

Algoritmo 4 Funções auxiliares para a identificação das agarras

L: uma lista de agarras
H: lista de agarras detectadas
W: lista de agarras candidatas em proporções do mundo real
l: cardinalidade de H e W (assume-se que possuam a mesma cardinalidade)

function ESCALA(H, W)
  $escala \leftarrow 0$
  for $h_i \in H$, $w_i \in W$, $i \in [1, l[$ **do**
    $distanciaReal \leftarrow$ DISTANCIA(w_i, w_{i-1})
    $distanciaImagem \leftarrow$ DISTANCIA(h_i, h_{i-1})
    $escala \leftarrow escala + (distanciaReal/distanciaImagem)$
  return $escala/(l-1)$

function DESLOCAPARAORIGEM($L, escala$)
  $referencia \leftarrow h_0$
  for $h_i \in L$, $i \in [0, l[$ **do**
    $h_i \leftarrow (referencia - h_i) * escala$

function ERRO(H, W)
  $maiorErro \leftarrow maxInt$
  for $h_i \in H$, $w_i \in W$, $i \in [0, l[$ **do**
    $distancia \leftarrow$ DISTANCIA(h_i, w_i)
    if $distancia > maiorErro$ **then**
      $maiorErro \leftarrow distancia$
  return $maiorErro$

function ERROCOMBINACAO(H, W)
  $escala \leftarrow$ ESCALA(H, M)
  $H \leftarrow$ DESLOCAPARAORIGEM($H, escala$)
  $W \leftarrow$ DESLOCAPARAORIGEM($W, 1$)
  return ERRO(H, W)

4.4.3 Estimativa da posição dos escaladores

Após a detecção do centro da massa dos escaladores e da identificação das agarras, o próximo passo é estimar a posição dos atletas no mundo real. Para tanto, subtrai-se as coordenadas do atleta das coordenadas de uma determinada agarra na imagem, o resultado é então multiplicado pela escala. Esse processo é realizado para cada agarra e a posição média é estimada. O algoritmo 5 descreve em detalhe os passos. A função PONTOMEDIO utilizada no algoritmo a seguir recebe como parâmetro uma lista de pontos e retorna o ponto médio.

Algoritmo 5 Estimativa da posição do escalador na parede

R: lista das agarras detectadas já identificadas
r_i^d: coordenadas de uma agarra na imagem
r_i^w: coordenadas de uma agarra no mundo real
c: coordenadas do centro de massa do escalador
$escala$: escala para transformar as medidas da imagem em medidas reais

$posicoes \leftarrow []$
for $r_i \in R$ **do**
    $diferencaPosicaoImagem \leftarrow c - r_i^d$
    $diferencaPosicaoReal \leftarrow diferencaPosicaoImagem * escala$
    $posicaoParede \leftarrow r_i^w + distanciaEscalaReal$
    $posicoes \leftarrow posicoes \cup [posicaoParede]$
return PontoMedio($posicoes$)

4.4.4 Pós-processamento

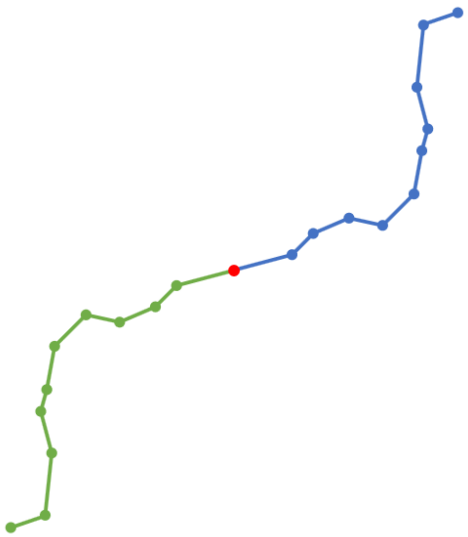
Após estimar a posição dos escaladores *frame* a *frame*, o passo seguinte consiste em corrigir a trajetória nos casos em que o mapeamento foi realizado de forma incorreta ou os casos em que a detecção dos atletas ou das agarras não funcionou. Para tanto, foi utilizado um filtro de mediana móvel de tamanho 10. Além disso, para que não houvesse perda dos dados iniciais e finais devido à aplicação do filtro, foram adicionados 50 pontos espelhados diagonalmente no começo e no fim da trajetória, de acordo com Smith (1989) esse procedimento apresenta melhores resultados se comparado com a simples replicação dos pontos das extremidades. A figura 28 mostra um exemplo de espelhamento e a figura 29 mostra um exemplo de aplicação do filtro.

4.5 Avaliação

A primeira etapa da avaliação contempla a análise de desempenho da rede *YOLO* ao detectar as agarras. Ao final do treinamento da rede, foram obtidas as métricas precisão, revocação, medida F e intersecção sobre união (*Intersection Over Union* ou IoU). Seguindo as definições apresentadas em Padilla, Netto e Silva (2020), cada uma delas será descrita à seguir.

Sendo Vp verdadeiro positivo, Fp falso positivo e Fn falso negativo, "a precisão de um modelo consiste na capacidade de identificar somente os objetos relevantes" (PADILLA;

Figura 28 – Exemplo de propagação dos dados para aplicação do filtro: em azul os pontos detectados, em verde os pontos propagados e em vermelho o ponto de referência utilizado para o espelhamento dos dados. Após a aplicação do filtro, os pontos espelhados são removidos.



Fonte – Daniel Freire Tsuha, 2023

NETTO; SILVA, 2020). Ou seja, o quão bom o modelo é em evitar marcações incorretas. A fórmula para se obter a precisão é dada por $P = Vp/(Vp + Fp)$.

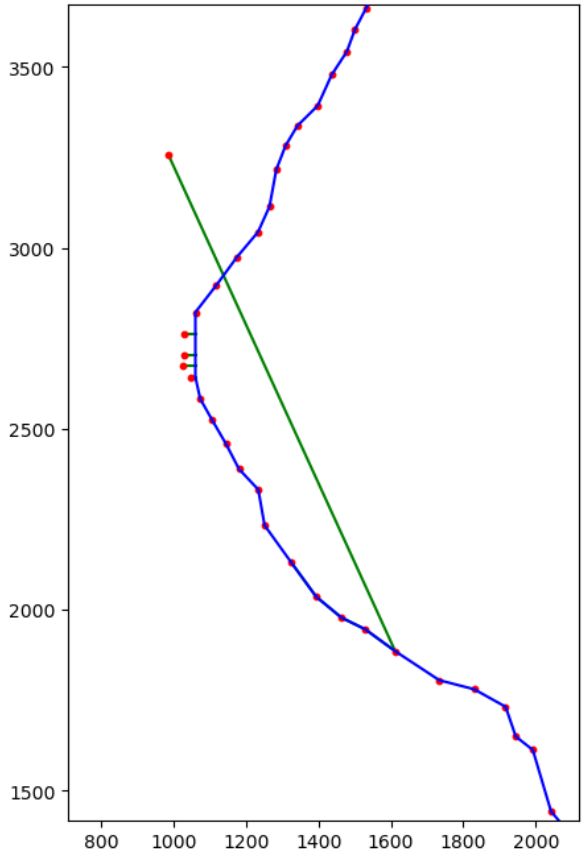
A revocação é definida como "a capacidade de um modelo de achar os casos relevantes" (PADILLA; NETTO; SILVA, 2020). Em outras palavras, de todas as agarras presentes na imagem, quantas agarras a rede foi capaz de detectar. O cálculo da revocação é dado por $R = Vp/(Vp + Fn)$.

Para se obter o valor da Medida F, calcula-se a média harmônica entre precisão e revocação com o objetivo de combiná-las. Quanto mais próximo de 1, melhor a capacidade da rede de detectar objetos (PADILLA; NETTO; SILVA, 2020). Calcula-se a medida F a partir da fórmula $F = 2(P * R)/(P + R)$.

No caso em que a rede é treinada para detectar objetos, a intersecção sobre união é uma medida que indica o quão precisa são as marcações realizadas pelo modelo (PADILLA; NETTO; SILVA, 2020). Para cada agarra detectada, compara-se a região obtida pela rede com a região rotulada do conjunto de treinamento, quanto maior a sobreposição das imagens, mais precisa é a rede. A figura 30 mostra como o cálculo é realizado.

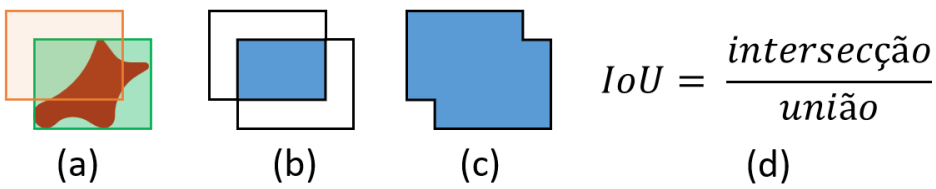
O processo para avaliar a precisão do algoritmo como um todo, consiste em comparar as posições obtidas de forma automática com as marcações realizadas manualmente. Para tanto, um *frame* válido de cada vídeo foi sorteado e o centro das agarras e o centro de

Figura 29 – Exemplo de aplicação do filtro de mediana móvel: em vermelho os pontos detectados pelo algoritmo, em azul a trajetória do atleta após a aplicação do filtro e em verde as distancias entre os pontos antes e depois da filtragem. Nos casos em que a posição do atleta foi estimada de forma errônea, o filtro é capaz de eliminar tais pontos e corrigir a trajetória.



Fonte – Daniel Freire Tsuha, 2023

Figura 30 – A intersecção sobre união tem como objetivo mensurar a precisão da rede ao gerar as coordenadas dos retângulos contento as agarras: a) em verde a região rotulada do conjunto de treinamento e em laranja a região detectada pela rede; b) intersecção das regiões; c) união das regiões; d) a métrica é calculada a partir da divisão das áreas de intersecção sobre união, quanto mais próxima de 1, mais preciso é o modelo.



Fonte – Daniel Freire Tsuha, 2023

massa aproximado dos atletas foram manualmente marcados. Para que um *frame* fosse considerado válido, ambos os atletas deveriam estar visíveis e a imagem não poderia retratar um momento após o término da prova. Além disso, não foram considerados os *frames* dos momentos em que houve muitas falhas consecutivas de mapeamento, já que tais trechos não puderam ser utilizados para reconstruir a trajetória do atleta.

Para validar a posição de uma agarra, considerou-se a posição do parafuso de fixação central. Baseados no gabarito apresentado na seção 2.1 que descreve as medidas da parede, tais pontos são os melhores candidatos para se estimar a posição das agarras. Para a marcação do centro de massa dos atletas foi considerado a região do cóccix. Após a marcação dos pontos, os dados obtidos foram processados para identificar as agarras (algoritmo 3) e para estimar a posição dos escaladores (algoritmo 5). Então se calculou o erro médio e o desvio padrão para as amostras.

As marcações manuais foram realizadas por um único operador, e a comparação foi feita considerando as distâncias verticais, horizontais e euclidianas das marcações realizadas pelo operador e as marcações realizadas de forma automática pelo algoritmo desenvolvido.

5 Resultados

O presente capítulo tem por objetivo apresentar os resultados obtidos após a execução do algoritmo descrito no capítulo anterior (capítulo 4). Os resultados foram divididos em três seções: a seção 5.1 apresenta os resultados do treinamento da rede responsável por detectar as agarras, já a seção 5.2 descreve o desempenho da rede *OpenPose* em identificar o centro de massa dos atletas, e pôr fim a seção 5.3 mostra os resultados para a estimativa da posição dos escaladores.

5.1 Detecção das agarras

Como descrito na subseção 4.2.2, o primeiro passo para detectar as agarras foi treinar uma rede neural convolucional *YOLO* (subseção 2.5.1). O passo seguinte foi calcular as métricas descritas na seção 4.5. Considerando-se um limiar de confiança de 0,75 (a certeza do modelo de que a região detectada possui uma agarra) e um limiar de 0,5 para o IoU, obteve-se os valores de 0,92 para precisão, 0,96 para a revocação e 0,94 para a medida F. A tabela 10 mostra os detalhes dos resultados.

Tabela 10 – Resultados do treinamento da rede YOLO.

Métrica	**Valor**
Precisão	0,92
Revocação	0,96
Medida F	0,94
IoU Média	86,94%
Vp	1404
Fp	119
Fn	59

Fonte – Daniel Freire Tsuha, 2023

Como valor de referência para estimar a posição dos escaladores na parede, o parafuso central foi utilizado para indicar as coordenadas, entretanto a rede foi treinada para detectar regiões retangulares contendo agarras. O centro geométrico dos retângulos foi considerado como uma aproximação da posição dos parafusos de fixação. Para mensurar o erro dessa abordagem, 533 agarras marcadas manualmente foram comparadas com os resultados obtidos pela rede. No eixo x o erro médio calculado foi de $4,0 \pm 3,4$ px, já no eixo y o erro foi de $5,6 \pm 4,5$ px e a distância média foi de $8,0 \pm 4,0$ px. A tabela 11 mostra as estatísticas completas e a figura 31 mostra a dispersão dos erros.

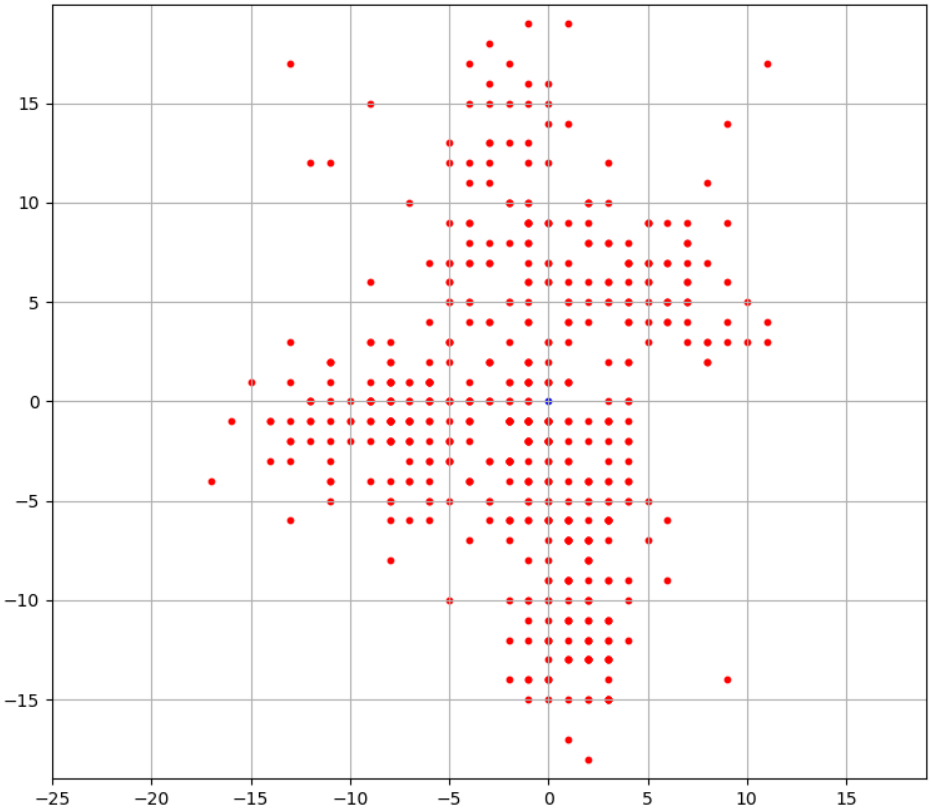
Apesar da medida em *pixels* não ser válida para estimar o erro no mundo real devido às diferentes escalas das imagens, os números demonstram que a aproximação utilizando regiões retangulares não afeta de forma drástica a estimativa da posição das agarras. Além disso, a posição de todas agarras detectadas são utilizadas no cálculo da escala da imagem, o que faz com que os erros sejam diluídos no processo, reduzindo a influência de agarras destoantes.

Tabela 11 – Erros de marcação do centro das agarras.

	Média	**Desvio padrão**	**Mínimo**	**Máximo**
Eixo x	4,0 px	± 3,4 px	0,0 px	17,0 px
Eixo y	5,6 px	± 4,5 px	0,0 px	19,0 px
Distância	8,0 px	± 4,0 px	1,0 px	21,4 px

Fonte – Daniel Freire Tsuha, 2023

Figura 31 – Dispersão dos erros da posição das agarras detectadas pela rede *YOLO* em *pixels*. A origem (ponto azul) indica a marcação manual e os pontos em vermelho as estimativas obtidas.



Fonte – Daniel Freire Tsuha, 2023

5.2 Detecção dos escaladores

A marcação manual dos centro de massa aproximado dos escaladores foi comparada com as marcações realizadas pela rede *OpenPose*. Para o cálculo do erro, dois *frames* foram descartados pois a rede não foi capaz de detectar corretamente um dos escaladores, sendo assim, o tamanho da amostra é de 158 escaladores. Os erros em *pixels* para amostra foram de $8,7 \pm 8,2$ px para o eixo x, $10 \pm 8,2$ px para o eixo y e o erro médio em termos de distância foi de $14,8 \pm 9,5$ px. A tabela 12 mostra todas as estatísticas calculadas.

Tabela 12 – Erros de marcação do centro de massa.

	Média	**Desvio padrão**	**Mínimo**	**Máximo**
Eixo x	8,7 px	± 8,2 px	0,0 px	42,2 px
Eixo y	10,0 px	± 8,2 px	0,1 px	50,8 px
Distância	14,8 px	± 9,5 px	0,4 px	52,2 px

Fonte – Daniel Freire Tsuha, 2023

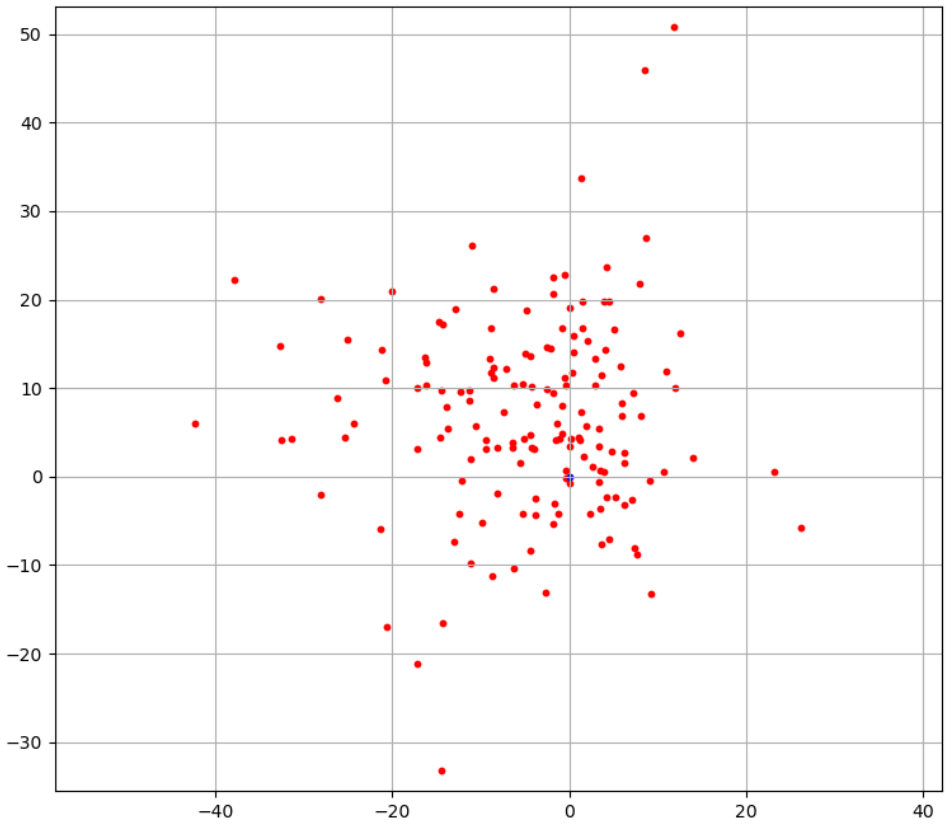
Ao comparar visualmente os resultados, é possível notar que a rede neural utilizada é consistente em detectar corretamente as articulações dos escaladores, mas apresenta variações quando há mudanças bruscas de ângulo ou pose. Nesses casos, o ponto usado como referência oscila de forma a representar a tridimensionalidade do corpo humano, principalmente quando a rede detecta lateralmente o quadril dos atletas. Além disso, a marcação do cóccix utilizada como referência para estimar o centro de massa não coincide exatamente com o ponto detectado pela rede.

Como cada vídeo apresenta uma escala métrica diferente devido à distância e angulação da câmera, o erro em *pixels* não influencia de forma equiparável todos os vídeos da amostra. Entretanto, o *OpenPose* se mostrou uma opção viável para estimar a posição do centro de massa do atletas na imagem. A figura 32 mostra a dispersão dos pontos ao redor do ponto marcado manualmente.

5.3 Estimativa de posição

A última etapa de avaliação do sistema consiste em comparar as estimativas das posições dos atletas obtidas a partir da marcação manual com as estimativas calculadas de forma automática pelo algoritmo desenvolvido. No eixo x o erro médio foi de $45,6 \pm 39,4$ mm, em y o erro foi de $58,7 \pm 77,3$ mm, em relação à distância dos pontos, o erro foi

Figura 32 – Dispersão dos erros da posição dos escaladores detectados pelo *OpenPose* em *pixels*. A origem (ponto azul) indica a marcação manual e os pontos em vermelho as estimativas obtidas.



Fonte – Daniel Freire Tsuha, 2023

de $82,2 \pm 79,3$ mm. A tabela 13 mostra as estatísticas completas e a figura 33 mostra a dispersão dos erros.

Tabela 13 – Erros de estimativa de posição.

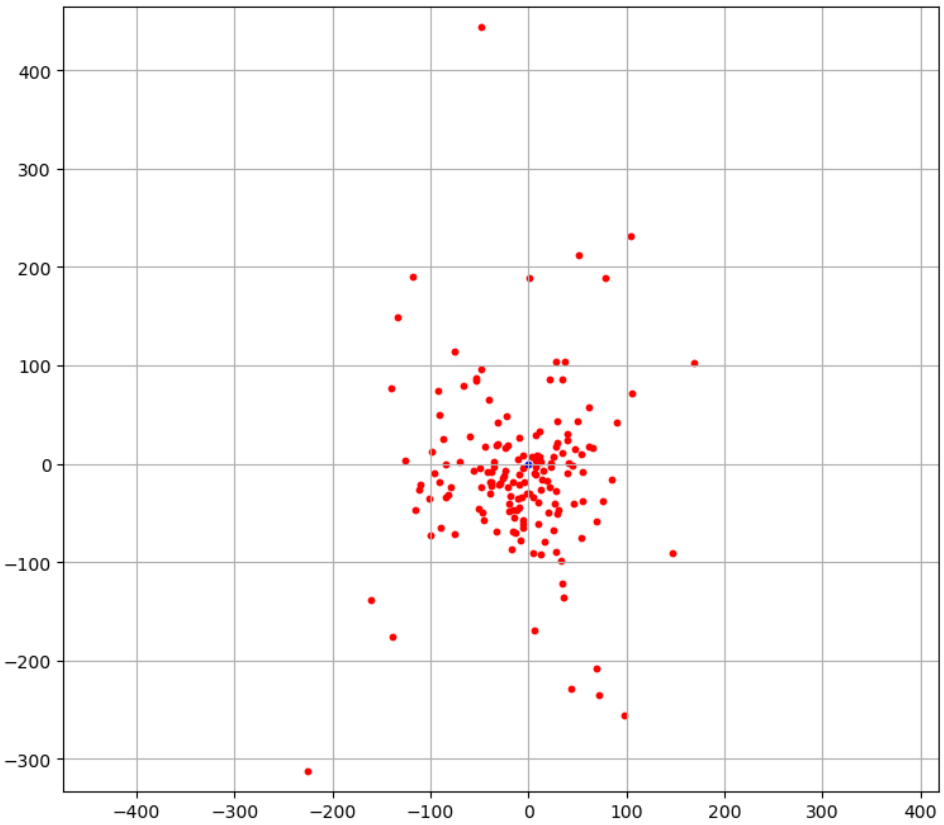
	Média	**Desvio padrão**	**Mínimo**	**Máximo**
Eixo x	45,4 mm	± 39,2 mm	1,19 mm	225,6 mm
Eixo y	55,5 mm	± 64,9 mm	0,36 mm	443,7 mm
Distância	79,2 mm	± 67,9 mm	7,24 mm	446,2 mm

Fonte – Daniel Freire Tsuha, 2023

Para efeitos de comparação, foi calculado o erro do algoritmo considerando as proporções da parede. Como descrito na seção 2.1, a parede possui 15 metros de altura por 3 de largura. No eixo x o erro médio foi de $1,51\% \pm 1,30\%$, em y o erro foi de $0,37\% \pm 0,43\%$. A tabela 14 mostra as estatísticas completas.

Ao analisar visualmente os casos em que as marcações apresentam as maiores discrepâncias, é possível observar que as principais causas de tais erros foram: problemas

Figura 33 – Dispersão dos erros das posições dos escaladores no mundo real em milímetros. A origem (ponto azul) indica a marcação manual e os pontos em vermelho as estimativas obtidas.



Fonte – Daniel Freire Tsuha, 2023

Tabela 14 – Erros de estimativa de posição em porcentagem.

	Média	**Desvio padrão**	**Mínimo**	**Máximo**
Eixo x	1,51%	± 1,30%	0,30%	7,52%
Eixo y	0,37%	± 0,43%	0,00%	2,95%

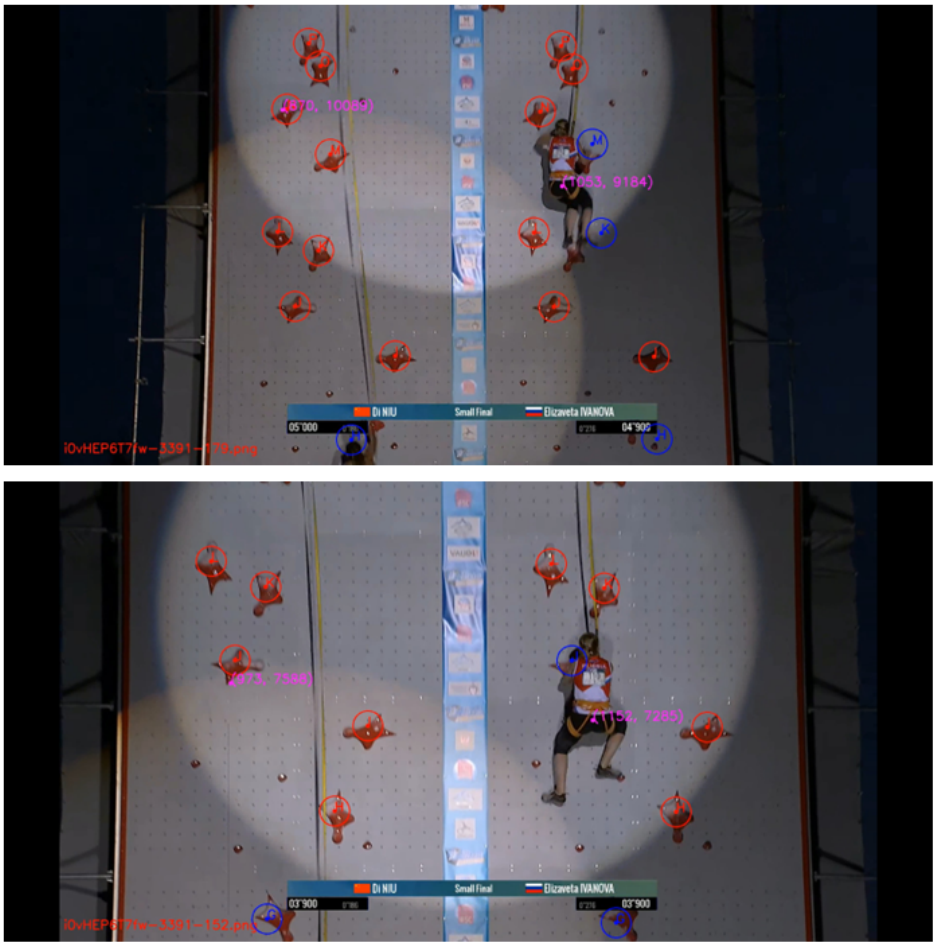
Fonte – Daniel Freire Tsuha, 2023

ao estimar a posição das agarras ocluídas, problemas ao mapear as agarras detectadas e problemas ao detectar a posição dos atletas. Cada um desses erros será discutido a seguir e como cada um deles influencia na precisão dos resultados.

O algoritmo que estima a posição das agarras ocluídas pelo corpo do escalador assume que a câmera se move exclusivamente na vertical e sem variações bruscas de ângulo ou de *zoom*, ou seja, que as imagens não sejam ampliadas ou reduzidas durante o decorrer da prova. Entretanto, em alguns vídeos essa premissa não se apresenta como verdadeira, em certos casos um dos atletas acaba por se atrasar e a distâncias entre os competidores aumenta. Para acomodar ambos os escaladores dentro da imagem, a câmera reduz o nível *zoom* fazendo com que o algoritmo não seja capaz de comparar a posição das agarras e

propagá-las. Essas variações acabam por gerar agarras em posições incorretas como mostra a figura 34.

Figura 34 – Erro na estimativa da posição de agarras ocluídas: as imagens mostram dois momentos distintos do mesmo vídeo com níveis de *zoom* diferentes. É possível notar que as agarras M e K foram propagadas de forma incorreta devido à mudança brusca da cena.



Fonte – International Federation of Sport Climbing

O mapeamento das agarras é a principal referência para estimar a posição dos atletas na parede. Apesar do algoritmo testar exaustivamente as combinações de agarras de forma a minimizar o erro, pressupõem-se que as agarras foram detectadas corretamente. Nesse sentido, as estimativas são sensíveis à agarras detectadas ou propagadas incorretamente, sendo essa a principal razão para a realização de mapeamentos incorretos. Mesmo quando a detecção das agarras é bem sucedida, ainda pode ocorrer variações na angulação da câmera em relação à parede, o que distorce a posição das agarras e consequentemente aferam a precisão do algoritmo. Tais erros podem impactar tanto na identificação das

agarras quanto no cálculo da posição dos atletas, reduzindo a precisão do algoritmo em estimar a posição dos atletas, a figura 35 exemplifica ambos os casos.

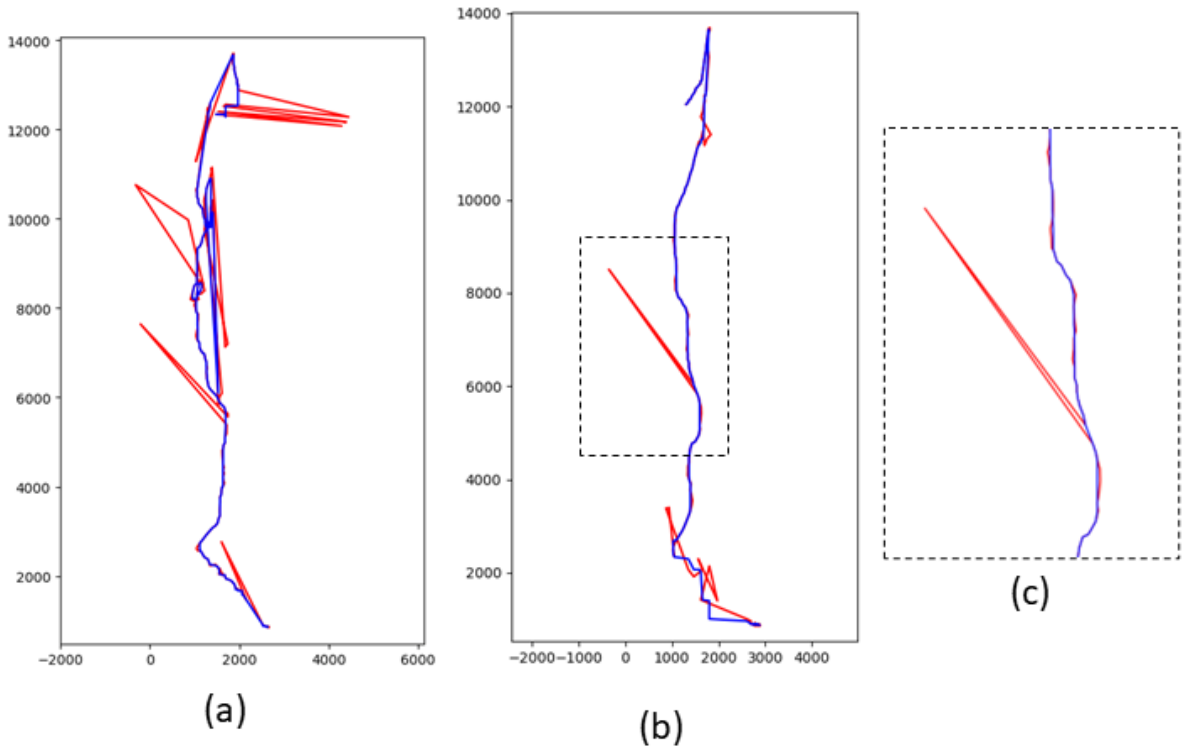
Figura 35 – Variação no ângulo da câmera: a linha traceja em vermelho mostra a influência da angulação da câmera em relação à parede. Essas variações impactam na precisão do algoritmo e ocorrem principalmente nos metros finais da prova. Em verde, um retângulo para referência visual caso a câmera estivesse perpendicular à parede.

Fonte – International Federation of Sport Climbing

O último caso de erro é quando a rede neural não é capaz de detectar corretamente a posição dos escaladores. Nesses casos, o *OpenPose* ou não foi capaz de detectar o atleta na imagem, ou o detectou erroneamente. Em todos os casos mencionados, o filtro de mediana foi aplicado para reduzir o efeito das falhas. Nos casos em que os erros ocorreram de forma isolada, o filtro foi capaz de corrigir as estimativas de posição. Entretanto, se os erros persistissem por uma longa sequência de *frames*, o filtro não dispunha de dados os suficientes para corrigir a trajetória. A figura 36 mostra os exemplos da aplicação do filtro em ambos os casos e a figura 37 compara as marcações manuais com as estimativas geradas pelo algoritmo.

Dados os pontos apresentados, a modificação ou inclusão de algumas etapas poderiam melhorar a precisão do algoritmo. Técnicas de homografia poderiam ser utilizadas para corrigir as distorções decorrentes da posição da câmera em relação à parede, principalmente nos metros finais da prova. Tais técnicas são utilizadas para alterar a perspectiva da imagem, nesse caso, fazendo que o ângulo de visão fosse perpendicular a parede de escalada. A inclusão dessa etapa no início do processo reduziria o erro de estimar a posição dos atletas.

Figura 36 – Aplicação do filtro de mediana móvel, em vermelho a trajetória original e em azul o resultado da aplicação do filtro: a) em alguns casos, o filtro não é capaz de reconstruir a trajetória de forma satisfatória, nesses casos pode ter ocorrido diversas falhas de detecção do algoritmo ou o atleta não estava visível em algum momento da prova; b) exemplos de em que o filtro foi bem sucedido em remover ruídos de estimativa de posição; c) ampliação da região em que o filtro foi bem sucedido em remover um ponto com erro.



Fonte – Daniel Freire Tsuha, 2023

Outra otimização que poderia ser feita é o recorte da região de interesse e redimensionamento das imagens de modo que a parede sempre tenha a mesma largura em *pixels*. Desse modo, o algoritmo que calcula a posição das agarras ocluídas não seria afetado por variações de *zoom*, o que geraria menos distorções entre os *frames* e, consequentemente, uma maior precisão ao calcular a distância entre as agarras. Uma outra otimização possível seria comparar a posição das agarras de ambas as paredes, tanto para estimar a posição de agarras ocluídas quanto para confirmar se o mapeamento das agarras foi realizado com sucesso.

Apesar das dificuldades listadas, o algoritmo apresentado obteve um erro médio inferior à seis centímetros no eixo y, que indica a altitude do atleta. Considerando que a parede possui 15 metros de altura, mesmo no pior caso, em que o erro foi de 44,3 centímetros, tal valor representa menos de três por cento da altura total da parede. Sendo assim, os

Figura 37 – Resultado final: rosa) marcação manual do parafuso de fixação da agarra; amarelo) marcação manual do centro de massa do escalador; verde) marcação automática das agarras; azul) marcação automática do centro de massa dos atletas.

Fonte – International Federation of Sport Climbing

resultados demonstram que é viável estimar a posição de atletas de forma automatizada a partir do processamento das gravações das competições de escalada esportiva.

6 Conclusões e trabalhos futuros

O presente trabalho apresentou uma forma automatizada de estimar a posição dos atletas durante as provas de escalada de velocidade utilizando uma combinação de técnicas computacionais para processar as gravações das competições. O algoritmo utiliza duas redes neurais convolucionais distintas para identificar os atletas e as agarras. Após a obtenção de tais dados, diversos passos são executados para combinar tais informações e estimar a posição dos atletas no decorrer da prova.

A primeira contribuição dessa pesquisa foi a realização de uma revisão sistemática da literatura sobre a utilização de técnicas de visão computacional para extrair informações de vídeos e avaliar o desempenho de atletas em diversas modalidades. A revisão cobriu diversos esportes como corrida, esportes aquáticos e esportes de quadra e listou as principais métricas de interesse de cada um deles. Além disso todas as técnicas computacionais aplicadas foram categorizadas, assim como as formas de avaliação e as amostras utilizadas. No momento da conclusão da revisão sistemática, não foram encontrados artigos que utilizassem técnicas de visão computacional aplicados à escalada esportiva.

A criação de um base de dados de vídeos rotulados também pode ser listada como uma das contribuições do presente trabalho, sendo que a base total contém 1515 agarras rotuladas. Tais marcações podem ser utilizadas para o treinamento de outras redes, sendo facilmente reaproveitados em outros projetos. Além disso, os pesos sinápticos resultantes do treinamento da rede *YOLO* também podem ser reutilizados, sendo um ponto de partida para outras pesquisas.

Do ponto de vista computacional, a contribuição dessa pesquisa consiste na criação e validação de um algoritmo capaz de reconstruir a trajetória dos atletas de escalada esportiva de forma automática e por meio da combinação de diversas técnicas computacionais. Em pesquisas futuras, espera-se que outras métricas como velocidade, aceleração, análise de movimento e comparação entre atletas possam ser extraídas a partir da ferramenta desenvolvida. O algoritmo mostrou ser capaz de realizar tais medições com um erro vertical inferior a 50 centímetros no pior caso avaliado, o que representa menos de 4% da altura total da parede. O trabalho também demonstrou que a utilização de redes neurais artificiais são uma alternativa de baixo custo e acessível para a análise de escaladores de velocidade.

Como contribuições para a ciência do esporte pode-se citar a redução do tempo para se analisar um vídeo de escalada de velocidade, já que a marcação e análise manual dos pontos de interesse é uma prática comum nessa área do conhecimento. Outra contribuição é a redução no custo do processo, já que nenhum equipamento específico é necessário para a obtenção dos resultados, já que a análise foi realizada em vídeos de competições já disponíveis. Além disso, espera-se que os profissionais da área possam se beneficiar das análises resultantes para otimizar o treino e detectar os pontos de melhoria dos atletas.

Por fim, a pesquisa mostrou ser possível utilizar técnicas de visão computacional para reconstruir a trajetória de atletas de escalada durante as competições de velocidade. Espera-se que esse seja o passo inicial para o desenvolvimento de ferramentas de baixo custo e que utilizem dados já disponíveis para analisar e auxiliar no treinamento e no desenvolvimento dos atletas de escalada esportiva em todo o mundo.

Como trabalhos futuros, algumas melhorias podem ser aplicadas para aumentar a precisão do algoritmo. Em relação as redes neurais convolucionais utilizadas, outros algoritmos como o *Region-based Convolutional Neural Networks* (R-CNN) poderiam ser testados para verificar se há aumento na precisão do detectores. No caso da detecção dos atletas, a utilização de outras abordagens para esqueletização dos atletas, como a *PoseNet*, poderiam ser utilizadas e comparadas com os resultados já obtidos. Já em relação à detecção das agarras, o algoritmo poderia ser treinado para gerar um polígono ao redor das agarras, e não um retângulo como foi utilizado na presente abordagem. Tal modificação tem o potencial de estimar com maior precisão a posição do parafuso central utilizado como referência.

Referências

ABREU, E. A. d. C.; ARAÚJO, S. R. S.; CANÇADO, G. H. d. C. P.; ANDRADE, A. G. P.; CHAGAS, M. H.; MENZEL, H.-J. K. Test-retest reliability of kinetic variables measured on campus board in sport climbers. *Sports Biomechanics*, Routledge, v. 18, n. 6, p. 649–662, 2019. Disponível em: ⟨https://doi.org/10.1080/14763141.2018.1456558⟩. Citado na página 15.

AGGARWAL, C. *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, 2018. ISBN 9783319944630. Disponível em: ⟨https://books.google.co.uk/books?id=achqDwAAQBAJ⟩. Citado 5 vezes nas páginas 28, 30, 31, 32 e 33.

AL-ALI, A.; ALMAADEED, S. A review on soccer player tracking techniques based on extracted features. In: *2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA)*. [S.l.: s.n.], 2017. p. 1–6. Citado na página 40.

ANDERSON, J. *An Introduction to Neural Networks*. MIT Press, 1995. (Bradford book). ISBN 9780262510813. Disponível em: ⟨https://books.google.co.uk/books?id=_ib4vPdB76gC⟩. Citado na página 28.

BAI, S.; KOLTER, J. Z.; KOLTUN, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *CoRR*, abs/1803.01271, 2018. Disponível em: ⟨http://arxiv.org/abs/1803.01271⟩. Citado na página 55.

BALLARD, D.; BROWN, C. *Computer Vision*. [S.l.]: Prentice-Hall, 1982. ISBN 9780131653160. Citado 2 vezes nas páginas 21 e 23.

BARONE, V.; VERDINI, F.; BURATTINI, L.; NARDO, F. D.; FIORETTI, S. A markerless system based on smartphones and webcam for the measure of step length, width and duration on treadmill. *Computer Methods and Programs in Biomedicine*, v. 125, p. 37–45, 2016. ISSN 0169-2607. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0169260715003260⟩. Citado 8 vezes nas páginas 44, 49, 50, 57, 58, 59, 60 e 61.

BARRIS, S.; BUTTON, C. A review of vision-based motion analysis in sport. *Sports Medicine*, v. 38, n. 12, p. 1025–1043, dez. 2008. ISSN 1179-2035. Disponível em: ⟨https://doi.org/10.2165/00007256-200838120-00006⟩. Citado 2 vezes nas páginas 38 e 39.

BEAUCHEMIN, S. S.; BARRON, J. L. The computation of optical flow. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 27, n. 3, p. 433–466, set. 1995. ISSN 0360-0300. Disponível em: ⟨http://doi.acm.org/10.1145/212094.212141⟩. Citado 2 vezes nas páginas 50 e 51.

BO, L.; SMINCHISESCU, C. Twin gaussian processes for structured prediction. *International Journal of Computer Vision*, v. 87, n. 1, p. 28, fev. 2009. ISSN 1573-1405. Disponível em: ⟨https://doi.org/10.1007/s11263-008-0204-y⟩. Citado na página 52.

BONACCORSO, G. *Machine Learning Algorithms*. Packt Publishing, 2017. ISBN 9781785884511. Disponível em: ⟨https://books.google.com.br/books?id=_-ZDDwAAQBAJ⟩. Citado 3 vezes nas páginas 24, 26 e 27.

BOONIM, K.; SANGUANSAT, P. Athletes performance evaluation by automated ladder for speed and agility. In: *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. [S.l.: s.n.], 2018. p. 106–109. Citado 7 vezes nas páginas 46, 47, 48, 49, 58, 59 e 60.

BURGER, W.; BURGE, M. *Digital Image Processing: An Algorithmic Introduction Using Java*. [S.l.]: Springer London, 2016. (Texts in Computer Science). ISBN 9781447166849. Citado 3 vezes nas páginas 21, 23 e 24.

CAO, Z.; HIDALGO, G.; SIMON, T.; WEI, S.; SHEIKH, Y. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *CoRR*, abs/1812.08008, 2018. Disponível em: ⟨http://arxiv.org/abs/1812.08008⟩. Citado na página 55.

CAO, Z.; MARTINEZ, G. H.; SIMON, T.; WEI, S.; SHEIKH, Y. A. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. Citado 3 vezes nas páginas 35, 36 e 68.

CESERACCIU, E.; SAWACHA, Z.; FANTOZZI, S.; CORTESI, M.; GATTA, G.; CORAZZA, S.; COBELLI, C. Markerless analysis of front crawl swimming. *Journal of Biomechanics*, v. 44, n. 12, p. 2236–2242, 2011. ISSN 0021-9290. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0021929011004362⟩. Citado 10 vezes nas páginas 14, 15, 45, 48, 52, 57, 58, 59, 60 e 61.

CHARMANT, J. *Kinovea*. 2017. Disponível em: ⟨https://www.kinovea.org/⟩. Citado 2 vezes nas páginas 15 e 38.

CHAUDHRY, R.; RAVICHANDRAN, A.; HAGER, G.; VIDAL, R. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2009. p. 1932–1939. Citado na página 51.

CHENG, H. D.; SHAN, J.; WANG, Y. Multiplayer tracking system for short track speed skating. *IET Computer Vision*, v. 8, n. 6, p. 629–641, dez. 2014. ISSN 1751-9632. Disponível em: ⟨http://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2014.0001⟩. Citado na página 14.

CHU, W.-T.; SITUMEANG, S. Badminton video analysis based on spatiotemporal and stroke features. In: *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*. New York, NY, USA: ACM, 2017. (ICMR '17), p. 448–451. ISBN 978-1-4503-4701-3. Disponível em: ⟨http://doi.acm.org/10.1145/3078971.3079032⟩. Citado 11 vezes nas páginas 45, 46, 48, 49, 50, 53, 57, 58, 59, 60 e 61.

COLYER, S. L.; EVANS, M.; COSKER, D. P.; SALO, A. I. T. A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. *Sports Medicine - Open*, v. 4, n. 1, p. 24, 2018. ISSN 2198-9761. Disponível em: ⟨https://doi.org/10.1186/s40798-018-0139-y⟩. Citado na página 40.

CORAZZA, S.; GAMBARETTO, E.; MUNDERMANN, L.; ANDRIACCHI, T. P. Automatic generation of a subject-specific model for accurate markerless motion capture and biomechanical applications. *IEEE Transactions on Biomedical Engineering*, v. 57, n. 4, p. 806–812, abr. 2010. Citado na página 53.

CORAZZA, S.; MUNDERMANN, L.; GAMBARETTO, E.; FERRIGNO, G.; ANDRIACCHI, T. P. Markerless motion capture through visual hull, articulated icp and subject specific model generation. *International Journal of Computer Vision*, v. 87, n. 1, p. 156, set. 2009. ISSN 1573-1405. Disponível em: ⟨https://doi.org/10.1007/s11263-009-0284-3⟩. Citado na página 52.

CRONIN, N. J.; RANTALAINEN, T.; AHTIAINEN, J. P.; HYNYNEN, E.; WALLER, B. Markerless 2d kinematic analysis of underwater running: A deep learning approach. *Journal of Biomechanics*, v. 87, p. 75–82, 2019. ISSN 0021-9290. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0021929019301551⟩. Citado 8 vezes nas páginas 45, 51, 55, 57, 58, 59, 60 e 61.

DAOUST, P. *Climbing has gone from niche sport to worldwide sensation. what is its dizzying appeal?* Guardian News and Media, 2018. Disponível em: ⟨https://www.theguardian.com/lifeandstyle/2018/aug/12/climbing-has-gone-from-niche-sport-to-worldwide-sensation-what-is-its-dizzying-appeal⟩. Citado na página 14.

DÍAZ-PEREIRA, M. P.; GóMEZ-CONDE, I.; ESCALONA, M.; OLIVIERI, D. N. Automatic recognition and scoring of olympic rhythmic gymnastic movements. *Human Movement Science*, v. 34, p. 63–80, 2014. ISSN 0167-9457. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0167945714000025⟩. Citado 6 vezes nas páginas 46, 54, 58, 59, 60 e 62.

DING, H.; CHENG, J.; LU, H.; ZHOU, Z. Synchronization analysis for synchronized diving videos. In: *2008 IEEE International Conference on Multimedia and Expo*. [S.l.: s.n.], 2008. p. 897–900. ISSN 1945-7871. Citado 8 vezes nas páginas 45, 48, 49, 50, 54, 58, 59 e 60.

EL-SALLAM, A. A.; BENNAMOUN, M.; SOHEL, F.; ALDERSON, J.; LYTTLE, A.; ROSSI, M. a. A low cost 3d markerless system for the reconstruction of athletic techniques. In: *2013 IEEE Workshop on Applications of Computer Vision (WACV)*. [S.l.: s.n.], 2013. p. 222–229. ISSN 1550-5790. Citado 10 vezes nas páginas 44, 47, 48, 49, 52, 57, 58, 59, 60 e 61.

EVANS, M.; COLYER, S.; COSKER, D.; SALO, A. Foot contact timings and step length for sprint training. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. [S.l.: s.n.], 2018. p. 1652–1660. Citado 8 vezes nas páginas 44, 49, 57, 58, 59, 60, 61 e 62.

FARNEBACK, G. Two-frame motion estimation based on polynomial expansion. In: BIGUN, J.; GUSTAVSSON, T. (Ed.). *Image Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003. p. 363–370. ISBN 978-3-540-45103-7. Citado na página 51.

FISCHER, M. T.; KEIM, D. A.; STEIN, M. Video-based analysis of soccer matches. In: *Proceedings Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports*. New York, NY, USA: Association for Computing

Machinery, 2019. (MMSports '19), p. 1–9. ISBN 9781450369114. Disponível em: ⟨https://doi.org/10.1145/3347318.3355515⟩. Citado na página 40.

FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, ACM, New York, NY, USA, v. 24, n. 6, p. 381–395, jun. 1981. ISSN 0001-0782. Disponível em: ⟨http://doi.acm.org/10.1145/358669.358692⟩. Citado na página 49.

FREUND, Y.; IYER, R.; SCHAPIRE, R. E.; SINGER, Y. An efficient boosting algorithm for combining preferences. *J. Mach. Learn. Res.*, JMLR.org, v. 4, p. 933–969, dez. 2003. ISSN 1532-4435. Disponível em: ⟨http://dl.acm.org/citation.cfm?id=945365.964285⟩. Citado na página 54.

GADE, R.; LARSEN, R. G.; MOESLUND, B. M. Measuring energy expenditure in sports by thermal video analysis. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. [S.l.: s.n.], 2017. p. 187–194. ISSN 2160-7516. Citado 11 vezes nas páginas 14, 42, 44, 48, 49, 57, 58, 59, 60, 61 e 62.

GAO, C.; CAI, Q.; MING, S. Yolov4 object detection algorithm with efficient channel attention mechanism. In: *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*. [S.l.: s.n.], 2020. p. 1764–1770. Citado 2 vezes nas páginas 35 e 67.

GASTEL, M. van; ZINGER, S.; KEMPS, H.; WITH, P. H. N. de. e-health video system for performance analysis in heart revalidation cycling. In: *2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)*. [S.l.: s.n.], 2014. p. 31–35. ISSN 2166-6814. Citado 6 vezes nas páginas 47, 48, 57, 58, 59 e 60.

GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2014. p. 580–587. ISSN 1063-6919. Citado na página 56.

GONZALEZ, R. C.; WOODS, R. E. *Processamento Digital de Imagens*. 3. ed. [S.l.]: Pearson, 2011. Citado 5 vezes nas páginas 21, 23, 48, 49 e 50.

GURNEY, K. *An Introduction to Neural Networks*. CRC Press, 2018. ISBN 9781482286991. Disponível em: ⟨https://books.google.co.uk/books?id=e0pZDwAAQBAJ⟩. Citado na página 28.

HORN, B. K.; SCHUNCK, B. G. Determining optical flow. *Artificial Intelligence*, v. 17, n. 1, p. 185–203, 1981. ISSN 0004-3702. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/0004370281900242⟩. Citado 2 vezes nas páginas 50 e 51.

JAIN, A. K. *Fundamentals of digital image processing*. [S.l.]: Englewood Cliffs, NJ: Prentice Hall,, 1989. Citado na página 38.

JIANG, P.; ERGU, D.; LIU, F.; CAI, Y.; MA, B. A review of yolo algorithm developments. *Procedia Computer Science*, v. 199, p. 1066–1073, 2022. ISSN 1877-0509. The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 2021): Developing Global Digital Economy after COVID-19. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/S1877050922001363⟩. Citado na página 34.

JINYAN, Y.; GUANLEI, X.; YU, L. Running state recognition in videos via frames' frequency and positions of two feet. In: *2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*. [S.l.: s.n.], 2013. p. 310–313. Citado 7 vezes nas páginas 14, 44, 57, 58, 59, 60 e 61.

JUNG, I.; SON, J.; BAEK, M.; HAN, B. Real-time mdnet. In: *Computer Vision – ECCV 2018*. Cham: Springer International Publishing, 2018. p. 89–104. ISBN 978-3-030-01225-0. Citado na página 56.

KASHIPAREKH, K.; NARWARIYA, J.; MALHOTRA, P.; VIG, L.; SHROFF, G. Convtimenet: A pre-trained deep convolutional neural network for time series classification. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2019. p. 1–8. Citado na página 67.

KELLEHER, J. *Deep Learning*. MIT Press, 2019. (The MIT Press Essential Knowledge series). ISBN 9780262537551. Disponível em: ⟨https://books.google.co.uk/books?id=b06qDwAAQBAJ⟩. Citado 5 vezes nas páginas 28, 30, 31, 32 e 33.

KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele Univ.*, v. 33, 8 2004. Citado 2 vezes nas páginas 40 e 63.

KOPOREC, G.; VUCKOVIC, G.; MILIC, R.; PERS, J. Quantitative contact-less estimation of energy expenditure from video and 3d imagery. *Sensors*, v. 18, n. 8, 2018. ISSN 1424-8220. Disponível em: ⟨https://www.mdpi.com/1424-8220/18/8/2435⟩. Citado 11 vezes nas páginas 44, 45, 46, 51, 54, 57, 58, 59, 60, 61 e 62.

KRUK, E. van der; REIJNE, M. M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European Journal of Sport Science*, Routledge, v. 18, n. 6, p. 806–819, 2018. PMID: 29741985. Disponível em: ⟨https://doi.org/10.1080/17461391.2018.1463397⟩. Citado na página 40.

KUBAT, M. *An Introduction to Machine Learning*. Springer International Publishing, 2017. ISBN 9783319639130. Disponível em: ⟨https://books.google.com.br/books?id=q6UzDwAAQBAJ⟩. Citado 4 vezes nas páginas 26, 27, 28 e 31.

LAFFAYE, G.; COLLIN, J.-M.; LEVERNIER, G.; PADULO, J. Upper-limb power test in rock-climbing. *Int J Sports Med*, v. 35, n. 8, p. 670–675, 2014. ISSN 0172-4622. 670. Disponível em: ⟨https://www.ncbi.nlm.nih.gov/pubmed/24554556⟩. Citado na página 15.

LEA, C.; FLYNN, M. D.; VIDAL, R.; REITER, A.; HAGER, G. D. Temporal convolutional networks for action segmentation and detection. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 1003–1012. ISSN 1063-6919. Citado na página 56.

LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, v. 1, n. 4, p. 541–551, 12 1989. ISSN 0899-7667. Disponível em: ⟨https://doi.org/10.1162/neco.1989.1.4.541⟩. Citado na página 31.

LEO, M.; D'Orazio, T.; TRIVEDI, M. A multi camera system for soccer player performance evaluation. In: *2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*. [S.l.: s.n.], 2009. p. 1–8. Citado 10 vezes nas páginas 45, 46, 48, 53, 54, 57, 58, 59, 60 e 61.

LI, B.; YAN, J.; WU, W.; ZHU, Z.; HU, X. High performance visual tracking with siamese region proposal network. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 8971–8980. ISSN 1063-6919. Citado na página 56.

LOWE, D. G. Object recognition from local scale-invariant features. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. [S.l.: s.n.], 1999. v. 2, p. 1150–1157 vol.2. Citado na página 50.

MANAFIFARD, M.; EBADI, H.; MOGHADDAM, H. A. A survey on player tracking in soccer videos. *Computer Vision and Image Understanding*, v. 159, p. 19–46, 2017. ISSN 1077-3142. Computer Vision in Sports. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S1077314217300309⟩. Citado na página 40.

MATAS, J. G.; GALAMBOS, C.; KITTLER, J. Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision and Image Understanding*, v. 78, n. 1, p. 119–137, 2000. ISSN 1077-3142. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S1077314299908317⟩. Citado na página 49.

MATHIS, A.; MAMIDANNA, P.; ABE, T.; CURY, K. M.; MURTHY, V. N.; MATHIS, M. W.; BETHGE, M. Markerless tracking of user-defined features with deep learning. *CoRR*, abs/1804.03142, 2018. Disponível em: ⟨http://arxiv.org/abs/1804.03142⟩. Citado na página 55.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, v. 5, n. 4, p. 115–133, Dec 1943. ISSN 1522-9602. Disponível em: ⟨https://doi.org/10.1007/BF02478259⟩. Citado na página 28.

MEHRIZI, R.; PENG, X.; XU, X.; ZHANG, S.; METAXAS, D.; LI, K. A computer vision based method for 3d posture estimation of symmetrical lifting. *Journal of Biomechanics*, v. 69, p. 40–46, 2018. ISSN 0021-9290. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0021929018300277⟩. Citado 10 vezes nas páginas 46, 47, 50, 52, 53, 58, 59, 60, 61 e 62.

MEHRIZI, R.; XU, X.; ZHANG, S.; PAVLOVIC, V.; METAXAS, D.; LI, K. Using a marker-less method for estimating l5/s1 moments during symmetrical lifting. *Applied Ergonomics*, v. 65, p. 541–550, 2017. ISSN 0003-6870. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0003687017300133⟩. Citado 9 vezes nas páginas 46, 47, 52, 53, 58, 59, 60, 61 e 62.

MICROSOFT. *Fazer Ajuste de Hiperparâmetro em um Modelo (V2) - Azure Machine Learning*. Microsoft, 2022. Disponível em: ⟨https://docs.microsoft.com/pt-br/azure/machine-learning/how-to-tune-hyperparameters⟩. Citado na página 27.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of Machine Learning*. MIT Press, 2012. (Adaptive Computation and Machine Learning series). ISBN 9780262304733. Disponível em: ⟨https://books.google.com.br/books?id=-ijiAgAAQBAJ⟩. Citado 4 vezes nas páginas 24, 25, 26 e 27.

MUKAI, R.; ASANO, T.; HARA, H. Analysis and evaluation of tennis plays by computer vision. In: *2011 IEEE International Conference on Mechatronics and Automation*. [S.l.:

s.n.], 2011. p. 784–788. ISSN 2152-744X. Citado 10 vezes nas páginas 14, 45, 46, 48, 54, 57, 58, 59, 60 e 61.

NAGANO, A.; FUJIMOTO, M.; KUDO, S.; AKAGUMA, R. An image-processing based technique to obtain instantaneous horizontal walking and running speed. *Gait & Posture*, v. 51, p. 7–9, 2017. ISSN 0966-6362. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0966636216305525⟩. Citado 7 vezes nas páginas 44, 48, 50, 57, 58, 59 e 60.

OLYMPIC. *IOC approves five new sports for Olympic Games Tokyo 2020*. 2016. Disponível em: ⟨https://www.olympic.org/news/ioc-approves-five-new-sports-for-olympic-games-tokyo-2020⟩. Citado na página 14.

ONG, A.; HARRIS, I. S.; HAMILL, J. The efficacy of a video-based marker-less tracking system for gait analysis. *Computer Methods in Biomechanics and Biomedical Engineering*, Taylor & Francis, v. 20, n. 10, p. 1089–1095, 2017. PMID: 28569549. Disponível em: ⟨https://doi.org/10.1080/10255842.2017.1334768⟩. Citado 5 vezes nas páginas 44, 57, 58, 59 e 60.

PADILLA, R.; NETTO, S. L.; SILVA, E. A. B. da. A survey on performance metrics for object-detection algorithms. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. [S.l.: s.n.], 2020. p. 237–242. ISSN 2157-8702. Citado 2 vezes nas páginas 76 e 77.

PANDUREVIC, D.; DRAGA, P.; SUTOR, A.; HOCHRADEL, K. Analysis of competition and training videos of speed climbing athletes using feature and human body keypoint detection algorithms. *Sensors*, v. 22, n. 6, 2022. ISSN 1424-8220. Disponível em: ⟨https://www.mdpi.com/1424-8220/22/6/2251⟩. Citado na página 16.

PARMAR, P.; MORRIS, B. T. Learning to score olympic events. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. [S.l.: s.n.], 2017. p. 76–84. ISSN 2160-7516. Citado 7 vezes nas páginas 45, 46, 54, 55, 58, 59 e 60.

PERS, J.; SULIC, V.; KRISTAN, M.; PERSE, M.; POLANEC, K.; KOVACIC, S. Histograms of optical flow for efficient representation of body motion. *Pattern Recognition Letters*, v. 31, n. 11, p. 1369–1376, 2010. ISSN 0167-8655. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0167865510001121⟩. Citado na página 51.

QIU, Z.; YAO, T.; MEI, T. Learning spatio-temporal representation with pseudo-3d residual networks. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017. p. 5534–5542. ISSN 2380-7504. Citado na página 56.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 779–788. ISSN 1063-6919. Citado na página 34.

REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE Computer Society, Los Alamitos, CA, USA, v. 39, n. 6, p. 1137–1149, jun. 2017. ISSN 1939-3539. Citado na página 56.

RIBEIRO, M. A. O.; NUNES, F. L. S. Left ventricle segmentation in cardiac mr: A systematic mapping of the past decade. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 54, n. 11s, sep 2022. ISSN 0360-0300. Disponível em: ⟨https://doi.org/10.1145/3517190⟩. Citado na página 38.

SANDLER, M.; HOWARD, A. G.; ZHU, M.; ZHMOGINOV, A.; CHEN, L. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *CoRR*, abs/1801.04381, 2018. Disponível em: ⟨http://arxiv.org/abs/1801.04381⟩. Citado na página 55.

SATO, S.; KOBAYASHI, N.; MIYASHITA, Y.; FUCHIDA, M.; NAKAMURA, A. Basic evaluation on soccer inside-kick proficiency. In: *2015 10th International Conference on Information, Communications and Signal Processing (ICICS)*. [S.l.: s.n.], 2015. p. 1–5. Citado 8 vezes nas páginas 45, 46, 50, 51, 54, 58, 59 e 60.

SHA, L.; LUCEY, P.; SRIDHARAN, S.; MORGAN, S.; PEASE, D. Understanding and analyzing a large collection of archived swimming videos. In: *IEEE Winter Conference on Applications of Computer Vision*. [S.l.: s.n.], 2014. p. 674–681. ISSN 1550-5790. Citado 9 vezes nas páginas 14, 45, 49, 50, 57, 58, 59, 60 e 61.

SHAO, L.; ZHU, F.; LI, X. Transfer learning for visual categorization: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, v. 26, n. 5, p. 1019–1034, 2015. Citado na página 67.

SHEETS, A. L.; ABRAMS, G. D.; CORAZZA, S.; SAFRAN, M. R.; ANDRIACCHI, T. P. Kinematics differences between the flat, kick, and slice serves measured using a markerless motion capture method. *Annals of Biomedical Engineering*, v. 39, n. 12, p. 3011, out. 2011. ISSN 1573-9686. Disponível em: ⟨https://doi.org/10.1007/s10439-011-0418-y⟩. Citado 8 vezes nas páginas 14, 45, 46, 48, 52, 58, 59 e 60.

SHIH, H.-C. A survey of content-aware video analysis for sports. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 28, n. 5, p. 1212–1231, maio 2018. Citado na página 39.

SHIN, J.; OZAWA, S. A study on motion analysis of an artistic gymnastics by using dynamic image processing - for a development of automatic scoring system of horizontal bar -. In: *2008 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2008. p. 1037–1042. ISBN 978-1-4244-2383-5. ISSN 1062-922X. Disponível em: ⟨http://ieeexplore.ieee.org/document/4811418/⟩. Citado 8 vezes nas páginas 14, 46, 48, 58, 59, 60, 61 e 62.

SIBELLA, F.; FROSIO, I.; SCHENA, F.; BORGHESE, N. A. 3d analysis of the body center of mass in rock climbing. *Human Movement Science*, v. 26, n. 6, p. 841–852, 2007. ISSN 1679457. Disponível em: ⟨https://www.sciencedirect.com/science/article/abs/pii/S0167945707000395⟩. Citado na página 14.

SIM, K. F.; SUNDARAJ, K. Human motion tracking on broadcast golf swing video using optical flow and template matching. In: *2010 International Conference on Computer Applications and Industrial Electronics*. IEEE, 2010. p. 169–173. ISBN 978-1-4244-9054-7. Disponível em: ⟨http://ieeexplore.ieee.org/document/5735069/⟩. Citado na página 14.

SKANSI, S. *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*. Springer International Publishing, 2018. (Undergraduate Topics in Computer Science). ISBN 9783319730042. Disponível em: ⟨https://books.google.com.my/books?id=5cNKDwAAQBAJ⟩. Citado 3 vezes nas páginas 28, 30 e 32.

SMITH, G. Padding point extrapolation techniques for the butterworth digital filter. *Journal of Biomechanics*, v. 22, n. 8, p. 967–971, 1989. ISSN 0021-9290. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/0021929089900821⟩. Citado na página 76.

SUARD, F.; RAKOTOMAMONJY, A.; BENSRHAIR, A.; BROGGI, A. Pedestrian detection using infrared images and histograms of oriented gradients. In: *2006 IEEE Intelligent Vehicles Symposium*. [S.l.: s.n.], 2006. p. 206–212. Citado na página 50.

SZELISKI, R. *Computer Vision: Algorithms and Applications*. 1st. ed. Berlin, Heidelberg: Springer-Verlag, 2010. ISBN 1848829345. Citado na página 38.

THOMAS, G.; GADE, R.; MOESLUND, T. B.; CARR, P.; HILTON, A. Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, v. 159, p. 3–18, 2017. ISSN 1077-3142. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S1077314217300711⟩. Citado na página 39.

TRAN, D.; BOURDEV, L. D.; FERGUS, R.; TORRESANI, L.; PALURI, M. Learning spatiotemporal features with 3d convolutional networks. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015. p. 4489–4497. Citado na página 55.

VEZZANI, R.; BALTIERI, D.; CUCCHIARA, R. People reidentification in surveillance and forensics: A survey. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 46, n. 2, dez. 2013. ISSN 0360-0300. Disponível em: ⟨https://doi.org/10.1145/2543581.2543596⟩. Citado na página 38.

VICTOR, B.; HE, Z.; MORGAN, S.; MINIUTTI, D. Continuous video to simple signals for swimming stroke detection with convolutional neural networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. [S.l.: s.n.], 2017. p. 122–131. ISSN 2160-7516. Citado 9 vezes nas páginas 14, 45, 48, 55, 57, 58, 59, 60 e 61.

WANG, H.; KLASER, A.; SCHMID, C.; LIU, C. Action recognition by dense trajectories. In: *CVPR 2011*. [S.l.: s.n.], 2011. p. 3169–3176. Citado 2 vezes nas páginas 50 e 51.

WANG, J.; QIU, K.; PENG, H.; FU, J.; ZHU, J. Ai coach: Deep human pose estimation and analysis for personalized athletic training assistance. In: *Proceedings of the 27th ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, 2019. (MM '19), p. 374–382. ISBN 9781450368896. Disponível em: ⟨https://doi.org/10.1145/3343031.3350910⟩. Citado 7 vezes nas páginas 46, 48, 53, 56, 57, 58 e 59.

WEI, S.; RAMAKRISHNA, V.; KANADE, T.; SHEIKH, Y. Convolutional pose machines. *CoRR*, abs/1602.00134, 2016. Disponível em: ⟨http://arxiv.org/abs/1602.00134⟩. Citado na página 55.

WEISSTEIN, E. *Geometric Centroid*. 2020. Disponível em: ⟨http://mathworld.wolfram.com/GeometricCentroid.html⟩. Citado na página 50.

WHITE, D. J.; OLSEN, P. D. A time motion analysis of bouldering style competitive rock climbing. *Journal of Strength and Conditioning Research*, v. 24, n. 5, p. 1356–1360, maio 2010. ISSN 1064-8011. Disponível em: ⟨https://insights.ovid.com/crossref?an=00124278-201005000-00028⟩. Citado na página 14.

WU, Y.; ZHAO, Z.; ZHANG, S.; YAO, L.; YANG, Y.; FU, T. Z. J.; WINKLER, S. Interactive multi-camera soccer video analysis system. In: *Proceedings of the 27th ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, 2019. (MM '19), p. 1047–1049. ISBN 9781450368896. Disponível em: ⟨https://doi.org/10.1145/3343031.3350586⟩. Citado 5 vezes nas páginas 45, 46, 58, 59 e 60.

XIANG, X.; TIAN, Y.; REITER, A.; HAGER, G. D.; TRAN, T. D. S3d: Stacking segmental p3d for action quality assessment. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2018. p. 928–932. ISSN 2381-8549. Citado 4 vezes nas páginas 56, 58, 59 e 60.

XIAO, B.; WU, H.; WEI, Y. Simple baselines for human pose estimation and tracking. In: FERRARI, V.; HEBERT, M.; SMINCHISESCU, C.; WEISS, Y. (Ed.). *Computer Vision – ECCV 2018*. Cham: Springer International Publishing, 2018. p. 472–487. ISBN 978-3-030-01231-1. Citado na página 57.

YAGI, K.; HASEGAWA, K.; SUGIURA, Y.; SAITO, H. Estimation of runners' number of steps, stride length and speed transition from video of a 100-meter race. In: *Proceedings of the 1st International Workshop on Multimedia Content Analysis in Sports*. New York, NY, USA: ACM, 2018. (MMSports'18), p. 87–95. ISBN 978-1-4503-5981-8. Disponível em: ⟨http://doi.acm.org/10.1145/3265845.3265850⟩. Citado 8 vezes nas páginas 44, 49, 55, 57, 58, 59, 60 e 61.

YENIKAYA, S.; YENIKAYA, G.; DUVEN, E. Keeping the vehicle on the road: A survey on on-road lane detection systems. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 46, n. 1, jul. 2013. ISSN 0360-0300. Disponível em: ⟨https://doi.org/10.1145/2522968.2522970⟩. Citado na página 38.

YU, Y.; LI, H.; YANG, X.; KONG, L.; LUO, X.; WONG, A. Y. L. An automatic and non-invasive physical fatigue assessment method for construction workers. *Automation in Construction*, v. 103, p. 1–12, 2019. ISSN 0926-5805. Disponível em: ⟨http://www.sciencedirect.com/science/article/pii/S0926580518308422⟩. Citado 8 vezes nas páginas 46, 47, 56, 57, 58, 59, 60 e 61.

ZECHA, D.; EGGERT, C.; EINFALT, M.; BREHM, S.; LIENHART, R. A convolutional sequence to sequence model for multimodal dynamics prediction in ski jumps. In: *Proceedings of the 1st International Workshop on Multimedia Content Analysis in Sports*. New York, NY, USA: ACM, 2018. (MMSports'18), p. 11–19. ISBN 978-1-4503-5981-8. Disponível em: ⟨http://doi.acm.org/10.1145/3265845.3265855⟩. Citado 7 vezes nas páginas 46, 55, 57, 58, 59, 60 e 61.

ZECHA, D.; LIENHART, R. Key-pose prediction in cyclic human motion. In: *2015 IEEE Winter Conference on Applications of Computer Vision*. [S.l.: s.n.], 2015. p. 86–93. ISSN 1550-5790. Citado 9 vezes nas páginas 45, 50, 53, 54, 57, 58, 59, 60 e 61.

ZHANG, Z.; PENG, H. Deeper and wider siamese networks for real-time visual tracking. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2019. p. 4586–4595. ISSN 1063-6919. Citado na página 56.

ZHOU, X.; HUANG, Q.; SUN, X.; XUE, X.; WEI, Y. Weakly-supervised transfer for 3d human pose estimation in the wild. *CoRR*, abs/1704.02447, 2017. Disponível em: ⟨http://arxiv.org/abs/1704.02447⟩. Citado na página 56.