

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO DE JOINVILLE
ENGENHARIA MECATRÔNICA

ARTHUR HAICKEL NINA

AVALIAÇÃO DE REDES NEURAIS PRÉ-TREINADAS PARA TAREFAS DE
ESTIMATIVA DE POSTURA HUMANA

Joinville
2024

ARTHUR HAICKEL NINA

AVALIAÇÃO DE REDES NEURAIS PRÉ-TREINADAS PARA TAREFAS DE
ESTIMATIVA DE POSTURA HUMANA

Trabalho apresentado como requisito parcial para obtenção do título de bacharel em Engenharia Mecatrônica, no Curso de Engenharia Mecatrônica, do Centro Tecnológico de Joinville, da Universidade Federal de Santa Catarina.

Orientador: Lucas Weihmann

Joinville
2024

Este trabalho é dedicado à minha família e amigos, que desde o princípio me incentivaram e me inspiraram a nunca desistir.

AGRADECIMENTOS

Em primeiro lugar, ao professor Lucas Weihmann, meu orientador, mentor e amigo, que propôs o tema que tanto me cativou desde o início. Sem seus direcionamentos, seria impossível seguir com o desenvolvimento e escrita deste trabalho.

Aos membros da banca, professores Ricardo e Pablo, que além de avaliadores deste trabalho, também foram mentores ao longo da minha vida acadêmica.

Aos membros da equipe Botcem e o professor Dalton, que atualmente a coordena, a quem devo tudo, por terem me ensinado a ser um bom profissional, um bom acadêmico e principalmente um líder.

Aos meus pais Vinicius e Rachel por terem sido meu alicerce nos momentos mais difíceis e compartilharem a alegria nos momentos mais felizes.

Aos meus amigos Pedro, Luís Filipe, Bruno, Aile, Maria Julia, Hannah, Luana por estarem sempre comigo independente de distância e tempo.

À minha família, na figura de Jéssica, Marianilce, Claudia, Miguel, Tânia e Eduardo por todo o amor e suporte.

Às famílias Dantas, Telles e Suguiura, por me receberem em seus lares como seu próprio filho, a que serei eternamente grato.

À minha psicóloga Sibeles e minha psiquiatra Emanuelle, que me ofereceram todo o suporte quanto à minha saúde mental, decisivo para que continuasse desenvolvendo este trabalho.

Aos colegas e amigos Allan, André, Lorena, Henrique e tantos outros que fiz durante essa caminhada em um ambiente, no início, totalmente novo.

E por fim, ao grupo Kokyu no Rolê, que me acompanha desde o início na mesma jornada acadêmica e que vou levar para a vida, pois se tornou a minha nova família.

*“ Seja grato por cada nascer do sol. Conquiste cada pôr do sol.
”*

John Cena

RESUMO

A prática de exercícios físicos é amplamente benéfica para a saúde física e mental. Contudo, movimentos incorretos ou realizados sem supervisão podem causar lesões e comprometer a qualidade de vida. Técnicas de visão computacional e inteligência artificial têm contribuído para democratizar o acesso a treinamentos supervisionados, com redes neurais de estimativa de poses humanas sendo amplamente distribuídas. Para isso, o estudo define um conjunto de métricas e conduz experimentos que permitem avaliar a eficácia dos modelos. Especificamente, os objetivos incluem selecionar os modelos a serem avaliados, determinar métricas de desempenho relevantes e validar a abordagem proposta por meio de estudos de caso em diferentes cenários. Este estudo busca comparar diferentes modelos pré-treinados, avaliando sua adequação para auxiliar atividades físicas em dispositivos com capacidade computacional limitada, analisando desempenho e identificando limitações. Os experimentos realizados demonstraram que o MoveNet apresentou o desempenho mais equilibrado em cenários desafiadores, mesmo em poses com articulações ocultas. Em comparação, o YOLO e o Mediapipe tiveram perdas significativas, incluindo falhas na detecção de keypoints e maiores erros máximos em métricas como o MPPE. Apesar disso, todos os modelos apresentaram limitações em ângulos desafiadores e poses complexas, indicando a necessidade de treinamentos mais diversificados e robustos. Os resultados ressaltam a importância de uma análise criteriosa de desempenho e destacam o MoveNet como o modelo mais consistente em geral. No entanto, melhorias adicionais são necessárias para garantir uma maior precisão e confiabilidade em aplicações práticas, especialmente em ambientes de baixa capacidade computacional.

Palavra-chave: Estimativa de postura humana. Inteligência artificial. Métricas.

ABSTRACT

The practice of physical exercise is widely beneficial for mental and physical health. However, incorrect movements or unsupervised activities can lead to injuries and reduced quality of life. Computer vision and artificial intelligence techniques have contributed to democratizing access to supervised training, with neural networks for human pose estimation being broadly available. This study defines a set of metrics and conducts experiments to evaluate the effectiveness of different pre-trained models. The objectives include selecting models, defining relevant performance metrics, and validating the proposed approach through case studies in diverse scenarios. The study focuses on assessing the models' suitability for assisting physical activities in low-computational environments, analyzing performance, and identifying limitations. Experiments revealed that MoveNet delivered the most balanced performance in challenging scenarios, even for poses with occluded joints. In contrast, YOLO and Mediapipe showed significant shortcomings, including keypoint detection failures and higher maximum errors in metrics like MPPE. Despite this, all models demonstrated limitations in challenging angles and complex poses, highlighting the need for more robust and diverse training. The results emphasize the importance of detailed performance analysis and identify MoveNet as the most consistent model overall. However, further improvements are essential to enhance accuracy and reliability in practical applications, especially in low-resource environments.

Keywords: Human pose estimation. Artificial intelligence. Metrics.

LISTA DE FIGURAS

Figura 1 – Modelos de corpos mais utilizados. (a) Modelo Esqueleto, (b) Modelo de Contorno, (c) Modelo de Volume.	16
Figura 2 – Diagrama de 17 keypoints	17
Figura 3 – Pipeline de rastreamento de poses humanas do Mediapipe.	18
Figura 4 – Diagrama de keypoints do Blazepose	19
Figura 5 – Arquitetura da rede Movenet	20
Figura 6 – Fluxograma do desenvolvimento do trabalho.	27
Figura 7 – Diagrama de pontos-chave selecionados no Roboflow.	28
Figura 8 – Anotação de keypoints sobre imagem no Roboflow.	28
Figura 9 – Exemplo de saída do código.	29
Figura 10 – Dataset com imagens de Yoga utilizado.	31
Figura 11 – Box plot do $Erro_{OKS}$ para o YOLO no estudo de caso 1.	32
Figura 12 – Perda de ponto-chave no calcanhar esquerdo pelo YOLO.	32
Figura 13 – Box plot do MPPE para o YOLO no estudo de caso 1.	33
Figura 14 – Pior caso do MPPE para o YOLO no estudo de caso 1	34
Figura 15 – Box plot do $Erro_{OKS}$ para o Mediapipe no estudo de caso 1.	34
Figura 16 – Perda de ponto-chave no calcanhar direito pelo Mediapipe.	35
Figura 17 – Box plot MPPE para o Mediapipe no estudo de caso 1.	36
Figura 18 – Pior caso apresentado pelo MPPE no Mediapipe.	36
Figura 19 – Box plot do $Erro_{OKS}$ para o Movenet no estudo de caso 1.	38
Figura 20 – Perda de ponto-chave no calcanhar direito pelo Movenet.	38
Figura 21 – Box plot do MPPE para o Movenet no estudo de caso 1.	39
Figura 22 – Erro MPPE destacado para o modelo Movenet em pose de yoga.	39
Figura 23 – Dataset com frames de exercícios de calistenia.	40
Figura 24 – Box plot do $Erro_{OKS}$ para o YOLO no estudo de caso 2.	41
Figura 25 – Ponto-chave da mão direita é perdido pelo YOLO.	41
Figura 26 – Ponto chave do nariz é perdido pelo YOLO.	42
Figura 27 – Box plot MPPE para o YOLO no estudo de caso 2.	42
Figura 28 – Box plot do $Erro_{OKS}$ para o Mediapipe no estudo de caso 2.	43
Figura 29 – Posicionamento incorreto de keypoints no braço direito, que está oculto.	44
Figura 30 – Box plot do MPPE para o Mediapipe no estudo de caso 2.	44
Figura 31 – Caso de outlier do MPPE no Mediapipe.	45
Figura 32 – Box plot do $Erro_{OKS}$ para o Movenet no estudo de caso 2.	45
Figura 33 – Posicionamento incorreto do nariz durante inferência do Movenet.	46

Figura 34 – Posicionamento incorreto do cotovelo direito durante inferência do Movenet.	46
Figura 35 – Box plot MPPE para o Movenet no estudo de caso 2.	47
Figura 36 – Médias para o $Erro_{OKS}$ no estudo de caso 1.	48
Figura 37 – Médias para MPPE no estudo de caso 1.	48
Figura 38 – Médias para o $Erro_{OKS}$ no estudo de caso 1.	49
Figura 39 – Médias para MPPE no estudo de caso 1.	49

LISTA DE QUADROS

Quadro 1 – Resultados de custo computacional sobre inferência em vídeo. . .	28
Quadro 2 – Maiores outliers apresentados em cada modelo para cada estudo de caso.	47

LISTA DE TABELAS

Tabela 1 – Valores de MPPE para o outlier da Figura 18.	37
Tabela 2 – Médias de OKS_{var} por ponto-chave.	54
Tabela 3 – Médias de $OKS_{0.1}$ por ponto-chave.	55

SUMÁRIO

1	INTRODUÇÃO	13
1.1	OBJETIVOS	14
1.1.1	Objetivo geral	14
1.1.2	Objetivos Específicos	14
1.2	ESTRUTURAÇÃO DO TEXTO	14
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	ESTIMATIVA DE POSTURA HUMANA	15
2.1.1	YOLO	16
2.1.2	Mediapipe	18
2.1.3	MoveNet	19
2.1.4	Outros modelos	21
2.2	MÉTRICAS DE AVALIAÇÃO DE ESTIMATIVA DE POSTURA	22
2.2.1	Object Keypoint Similarity	22
2.2.2	Erro relativo de dispersão OKS	22
2.2.3	Mean Per Part Error	23
2.3	TRABALHOS CORRELATOS	24
3	METODOLOGIA	26
4	EXPERIMENTOS E RESULTADOS	31
4.1	ESTUDO DE CASO 1	31
4.1.1	YOLO	31
4.1.2	Mediapipe	33
4.1.3	Movenet	37
4.2	ESTUDO DE CASO 2	40
4.2.1	YOLO	40
4.2.2	Mediapipe	43
4.2.3	Movenet	45
4.3	SÍNTESE DE DADOS	48
5	CONCLUSÃO	50
	REFERÊNCIAS	52
	APÊNDICE A – TABELA DE MÉDIAS DE OKS POR KEYPOINT POR MODELO PARA AS POSES DE YOGA	54
	APÊNDICE B – TABELA DE MÉDIAS DE OKS POR KEYPOINT POR MODELO PARA AS POSES DE YOGA	55

1 INTRODUÇÃO

É amplamente aceito que a prática de exercícios físicos traz benefícios significativos para a saúde motora, cardiovascular, mental e neurológica do ser humano. No entanto, a execução inadequada de movimentos, especialmente em atividades não supervisionadas, pode resultar em lesões graves, conforme detalhado por Souza *et al.* (2015).

Para mitigar esses riscos, técnicas de estimativa de poses humanas (HPE, do inglês Human Pose Estimation) — problema central da Visão Computacional, que visa localizar pontos-chave do corpo humano em imagens ou vídeos, representando articulações ou partes específicas (Josyula; Ostadabbas, 2021) — têm se mostrado promissoras no desenvolvimento de assistentes virtuais para monitoramento e correção de movimentos durante exercícios físicos. Modelos de pose estimation utilizam algoritmos de aprendizado profundo (deep learning) para identificar e analisar posições corporais, mas apresentam desafios quando aplicados em contextos reais, como ambientes móveis, onde há limitações de processamento, iluminação variável e ângulos de câmera diversos.

Embora existam diversos modelos pré-treinados e amplamente utilizados para estimativa de poses, como OpenPose, MediaPipe e YOLO-Pose, ainda não há consenso sobre uma métrica universalmente aceita para avaliar o desempenho desses modelos em tarefas específicas. Estudos existentes frequentemente adotam métricas como OKS (Object Keypoint Similarity) e MPPE (Mean Per Part Error), mas não oferecem uma comparação abrangente que considere aplicações práticas e condições não ideais. Dessa forma, há uma lacuna em estudos que avaliem e comparem sistematicamente a robustez desses modelos frente a situações desafiadoras, especialmente em dispositivos móveis.

Além disso, a popularização dos smartphones, com mais de 8 bilhões de dispositivos em uso no mundo atualmente Richter (2023), posiciona o mercado mobile como um dos principais canais de distribuição de software. Essa realidade torna imprescindível que ferramentas de assistência baseadas em estimativa de poses sejam acessíveis por meio de plataformas móveis, seja por aplicativos nativos ou páginas web.

Para esclarecer os objetivos e delimitações deste trabalho, é importante ressaltar que não será realizado treinamento de redes neurais ou o desenvolvimento de um novo modelo. O foco está na avaliação de modelos pré-treinados em seu estado padrão de distribuição, sem ajustes adicionais. Este trabalho não propõe uma nova rede neural, mas sim a sistematização de um procedimento de avaliação para esses modelos de estimativa de postura humana.

Dessa forma, busca-se avançar nesse campo ao propor uma análise sistemática de modelos de estimativa de poses humanas, com foco em sua aplicação na assistência de exercícios físicos em dispositivos móveis. Além de discutir as métricas utilizadas, o trabalho avalia o desempenho de diferentes modelos em cenários reais e propõe insights para melhorias futuras, considerando as limitações e desafios identificados.

1.1 OBJETIVOS

Para desenvolver uma metodologia sistemática de comparação de modelos de estimativa de postura humana, propõem-se os seguintes objetivos.

1.1.1 Objetivo geral

Realizar um comparativo, baseado em um conjunto de métricas e resultados, entre modelos pré-treinados de pose estimation para a tarefa de auxiliar com atividades físicas em ambientes de capacidade computacional limitada.

1.1.2 Objetivos Específicos

- escolher modelos de pose estimation a serem comparados
- definir um conjunto métricas para realizar o comparativo
- validar as métricas definidas por meio de um estudo de caso

1.2 ESTRUTURAÇÃO DO TEXTO

O trabalho está estruturado da seguinte forma: no Capítulo 2, são apresentados os fundamentos teóricos sobre estimativa de postura, os modelos considerados para avaliação e as métricas utilizadas no estudo de caso. No Capítulo 3, é detalhada a metodologia adotada para o desenvolvimento do projeto, seguida pela discussão dos resultados obtidos nos experimentos, com ênfase nos casos de maiores desvios, no Capítulo 4. Por fim, no Capítulo 5, é realizado um balanço comparativo entre os modelos testados e são apresentadas sugestões para trabalhos futuros que possam aprofundar a avaliação de redes neurais para estimativa de pose.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo detalha o conhecimento formal que embasa o trabalho, listando e apresentando os princípios de funcionamento das redes neurais exploradas e métricas utilizadas para o comparativo.

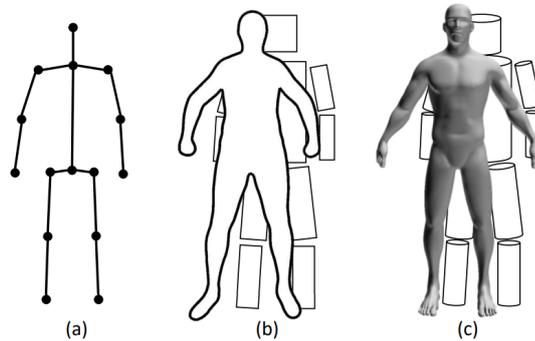
2.1 ESTIMATIVA DE POSTURA HUMANA

A estimativa de postura humana (HPE) é um problema central na Visão Computacional, cujo objetivo é determinar a posição e orientação de uma pessoa ou objeto em imagens ou vídeos. Geralmente, isso envolve a previsão da localização de pontos-chave (keypoints), que representam articulações ou partes específicas do corpo humano. Esses pontos podem ser modelados em 2D ou 3D, utilizando técnicas clássicas ou algoritmos baseados em aprendizado profundo. Atualmente, o aprendizado profundo é amplamente utilizado devido à capacidade das redes neurais de modelar relações complexas e lidar com condições desafiadoras, como oclusões ou ruídos ambientais. Esse processo é uma pedra fundamental na tecnologia de captura de movimentos, como afirma Rohan *et al.* (2020), sendo relevante em áreas como saúde, esportes, segurança e interação humano-computador (Badiola-Bengoia; Mendez-Zorrilla, 2021).

De acordo com Chen *et al.* (2020) estimativa de posturas utilizando aprendizado profundo pode ser classificada em diferentes abordagens. Métodos generativos utilizam modelos estruturais do corpo humano como base, enquanto métodos discriminativos operam diretamente sobre dados visuais. Na estimativa de múltiplas pessoas, dois paradigmas são comuns: top-down, que começa detectando indivíduos e aplicando estimativa de postura em cada um, e bottom-up, que identifica partes do corpo de todas as pessoas antes de agrupá-las com base em similaridades. Enquanto o método top-down apresenta aumento de custo computacional proporcional ao número de indivíduos, o bottom-up mantém a eficiência constante. Além disso, as abordagens podem ser de estágio único, mapeando imagens diretamente para posturas, ou multi-estágios, onde previsões são refinadas progressivamente.

Na Figura 1 é possível observar os modelos de corpos humanos mais comuns. Dentre estes há o modelo esqueleto, o modelo de contorno e o modelo de volume. O modelo esqueleto utiliza pontos-chave para representar articulações como tornozelos, joelhos e ombros, sendo amplamente aplicado em estimativas 2D e 3D devido à sua simplicidade e flexibilidade. Por outro lado, modelos de contorno usam formas geométricas como retângulos ou silhuetas para descrever as partes do corpo, enquanto modelos volumétricos representam o corpo com estruturas tridimensionais, como ci-

Figura 1 – Modelos de corpos mais utilizados. (a) Modelo Esqueleto, (b) Modelo de Contorno, (c) Modelo de Volume.



Fonte: Chen *et al.* (2020)

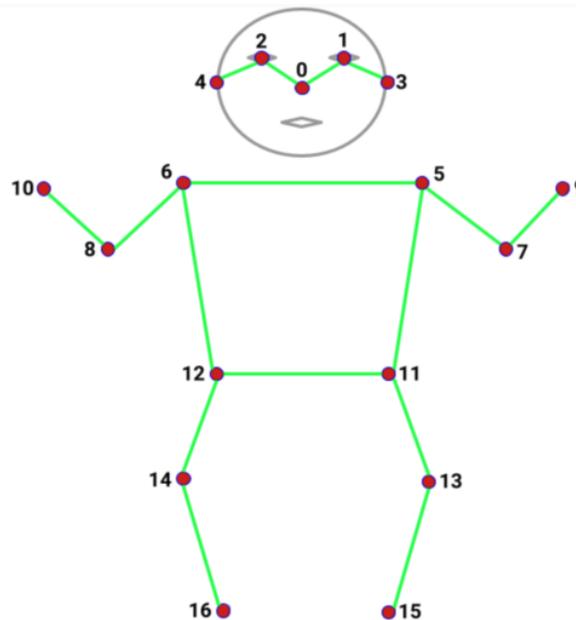
lindros, cones ou malhas gráficas detalhadas. Apesar de mais complexos, os modelos volumétricos mais recentes, baseados em escaneamento 3D, oferecem maior precisão e detalhamento, sendo ideais para aplicações que exigem representações realistas do corpo humano. Para este trabalho, foi selecionado o modelo esqueleto devido à sua ampla utilização e facilidade de implementação, características que o tornam adequado para atender aos objetivos propostos neste estudo de caso.

2.1.1 YOLO

O YOLO-Pose é uma extensão da arquitetura de detecção YOLOv8, projetada para realizar a estimativa de pose 2D de múltiplas pessoas em imagens de maneira eficiente e sem o uso de mapas de calor (heatmaps). Os heatmaps são representações bidimensionais utilizadas em redes neurais para localizar pontos-chave do corpo humano em imagens. Cada ponto-chave (como joelho ou pulso) é associado a um mapa de probabilidade, onde os valores mais altos indicam a localização provável da junta correspondente. A rede neural convolucional gera um heatmap para cada ponto-chave, e a posição final é determinada ao localizar o ponto de maior intensidade no mapa. (Dubey; Dixit, 2023) Apesar de eficaz, esse método introduz uma alta complexidade computacional e exige maior uso de memória, tornando-o menos eficiente para aplicações em dispositivos com recursos limitados.

Redes neurais convolucionais (CNN, do inglês Convolutional Neural Network) são modelos de aprendizado de máquina inspirados na estrutura do cérebro humano, projetados para processar e analisar dados em formato de grade, como imagens. Suas camadas convolucionais aplicam filtros para extrair características relevantes, como bordas, texturas e formas, reduzindo a dimensionalidade e preservando informações essenciais. Essa arquitetura é amplamente utilizada em tarefas de visão computacional, como reconhecimento de objetos e estimativa de pose humana, devido à

Figura 2 – Diagrama de 17 keypoints



Fonte: Tensorflow (2022)

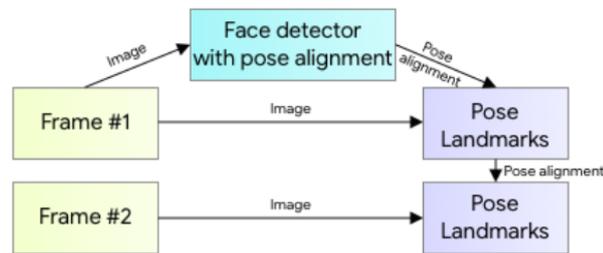
sua capacidade de capturar padrões espaciais em diferentes escalas. (Changotra *et al.*, 2022)

Diferentemente dessa abordagem, o YOLO-Pose elimina os heatmaps e prevê diretamente as coordenadas (x, y) dos pontos-chave. Essa adaptação visa superar as limitações de métodos tradicionais, que frequentemente requerem estágios separados de detecção e agrupamento de pontos-chave, gerando alta latência. No YOLO-Pose, os pontos-chave são associados diretamente às âncoras de detecção, permitindo que cada âncora armazene não apenas a bounding box (caixa delimitadora da pessoa detectada), mas também as coordenadas dos pontos-chave. Isso simplifica o processo, reduzindo a complexidade e o consumo de recursos computacionais, ao mesmo tempo em que possibilita uma estimativa de pose mais eficiente e adequada para cenários com restrições de hardware, como dispositivos móveis.

Essa abordagem heatmap-free permite uma integração mais direta com sistemas de visão computacional baseados em detecção de objetos, sem aumento significativo de custo computacional. O YOLO-Pose combina características de métodos top-down e bottom-up, oferecendo uma solução unificada com processamento em tempo constante e baixa latência, adequando-se, assim, a aplicações em tempo real de detecção e estimação de pose. (Maji *et al.*, 2022)

A arquitetura do YOLO-Pose é baseada no backbone CSP-Darknet53 e na rede PANet, que possibilitam a fusão de informações de múltiplas escalas, facilitando a detecção robusta e precisa de pontos-chave mesmo em cenários complexos com grande variação de escala e oclusões. A rede realiza duas frentes de detecção, sendo

Figura 3 – Pipeline de rastreamento de poses humanas do Mediapipe.



Fonte: Bazarevsky *et al.* (2020)

uma dedicada à previsão das caixas delimitadoras, que indicam a localização e o tamanho das pessoas detectadas na imagem, e a outra focada na previsão das coordenadas dos 17 pontos-chave (keypoints) que representam articulações ou partes específicas do corpo de cada pessoa detectada, como ilustra a Figura 2. Além disso, o YOLO-Pose utiliza a Object Keypoint Similarity (OKS), que será apresentada na seção 2.2.1, tanto como métrica de avaliação quanto como função de perda durante o treinamento, o que permite ao modelo dar maior peso a pontos-chave críticos, como aqueles da região da cabeça, otimizando diretamente o desempenho final em tarefas de pose humana.

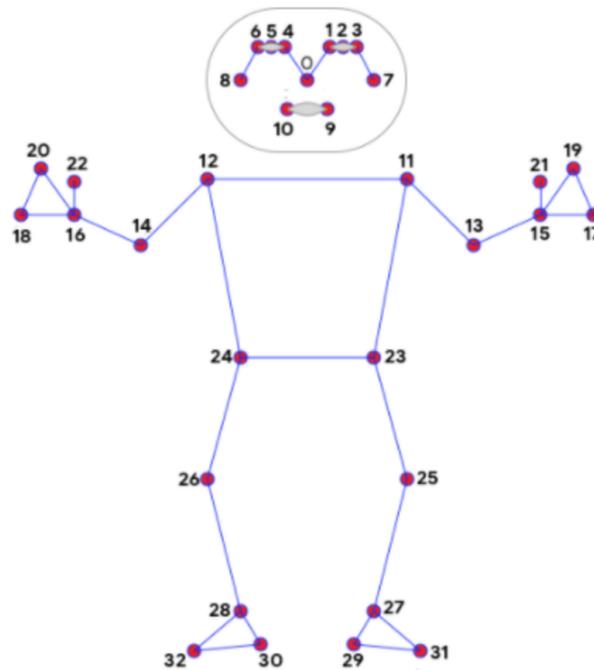
2.1.2 Mediapipe

O Mediapipe utiliza o modelo BlazePose, com técnicas avançadas de detecção e rastreamento de poses 3D humanas com uma arquitetura em dois estágios: detecção de pessoa e regressão de keypoints. Para isso, utiliza um detector rápido que localiza o rosto da pessoa, simplificando a tarefa de encontrar o torso e outras partes corporais, aproveitando a posição fixa da cabeça em muitas aplicações.

A Figura 3 ilustra o pipeline de detecção e rastreamento de poses humanas utilizado pelo MediaPipe. Inicialmente, um detector leve identifica os pontos-chave (landmarks) do rosto, que atuam como âncoras para alinhar os landmarks do torso e, subsequentemente, do restante do corpo. Essa abordagem evita o uso de técnicas mais complexas, como a supressão de máxima não-máxima (NMS), comumente empregadas em métodos de detecção de objetos para resolver problemas de previsões sobrepostas. O NMS funciona ao selecionar apenas a previsão com o maior score em uma determinada área e suprimir as demais, o que pode introduzir maior complexidade computacional. Ao priorizar keypoints faciais e corporais em uma sequência bem definida, o pipeline do MediaPipe simplifica o processo de detecção e mantém a eficiência, mesmo em dispositivos com restrições de hardware.

Após a detecção inicial, o BlazePose utiliza uma rede neural que combina heatmaps e regressão direta de coordenadas para prever 33 pontos-chave do corpo

Figura 4 – Diagrama de keypoints do Blazepose



Fonte: Bazarevsky *et al.* (2020)

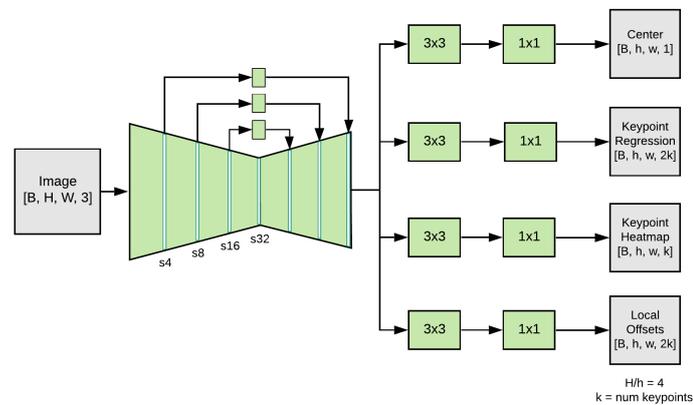
humano, vide Figura 4. Durante o treinamento, heatmaps supervisionados ajudam a otimizar o aprendizado, mas são descartados na inferência para simplificar o modelo, mantendo apenas a rede de regressão para prever diretamente as coordenadas dos pontos-chave.

A arquitetura encoder-decoder usa conexões diretas para preservar informações de baixo e alto nível, ajustando automaticamente a posição, rotação e escala do indivíduo em tempo real. Isso permite que o BlazePose funcione de maneira robusta, mesmo em cenários com oclusões e variações de pose. Técnicas de augmentation como simulação de oclusões e ajustes de escala tornam o modelo mais robusto para situações do mundo real, garantindo precisão e eficiência em dispositivos móveis. (Bazarevsky *et al.*, 2020)

2.1.3 MoveNet

Com o objetivo de facilitar a captura de posturas, será utilizado o modelo pré-treinado TensorFlow MoveNet. Esse modelo é capaz de identificar os mesmos 17 pontos da Figura 2, sendo um dos principais modelos do estado da arte para aplicações focadas em saúde e bem-estar, especialmente para reconhecer posturas em exercícios físicos. O modelo oferece duas variantes: Lightning e Thunder. A variante Lightning será a escolhida para este trabalho, pois é otimizada para aplicações que exigem baixa latência, como em dispositivos móveis ou na web. Já a variante Thunder é mais indicada para casos que demandam maior precisão e tem um maior custo

Figura 5 – Arquitetura da rede Movenet



Fonte: Votel e Li (2021)

computacional.

A arquitetura MoveNet combina um extrator de características MobileNetV2 com uma Feature Pyramid Network (FPN), criando um modelo eficiente e preciso para estimativa de poses humanas. A FPN é responsável por integrar informações de múltiplas escalas, permitindo a geração de mapas ricos em detalhes espaciais e semânticos com resolução refinada. Isso resulta em mapas de stride 4, que oferecem maior precisão na localização de keypoints.

O processo operacional do MoveNet ocorre em quatro etapas principais. Primeiro, é gerado um heatmap central que localiza o centro do corpo humano na imagem. Em seguida, um campo de regressão prevê os keypoints iniciais com base nas características detectadas. A terceira etapa utiliza heatmaps individuais para cada keypoint, ponderando a proximidade das articulações. Por fim, um campo de deslocamento 2D ajusta as coordenadas finais, melhorando a precisão dos keypoints preditos.

Essa combinação de componentes faz com que o MoveNet alcance alto desempenho em tarefas de estimativa de poses em tempo real, sendo robusto em cenários desafiadores como oclusões e variações de escala. Sua eficiência e escalabilidade tornam-no adequado para aplicações em dispositivos móveis e sistemas de visão embarcados. (Votel; Li, 2021)

O modelo recebe como entrada um frame de vídeo ou uma imagem, representado por um tensor do tipo int32, com tamanho $256 \times 256 \times 3$, contendo os valores RGB no intervalo de $[0; 255]$. A saída é um tensor do tipo float32, com dimensões $[1; 1; 17; 3]$, onde os dois primeiros canais da última dimensão representam as coordenadas x e y (normalizadas no intervalo $[0; 1]$). O terceiro canal indica a pontuação da predição de cada ponto-chave no intervalo $[0; 1]$, é possível observar este funcionamento na Figura 5.

2.1.4 Outros modelos

Durante a fase de pesquisa de redes neurais para executar os experimentos, outros modelos de HPE testados são apresentados brevemente nesta subsecção.

O OpenPose é um sistema em tempo real para estimativa de poses humanas em 2D, projetado para detectar múltiplas pessoas em uma única imagem ou vídeo. Ele utiliza uma abordagem bottom-up que identifica partes do corpo individualmente e, em seguida, associa essas partes aos respectivos indivíduos por meio de Part Affinity Fields (PAFs). Os PAFs são campos vetoriais 2D que codificam a orientação e a conexão entre partes do corpo, como ombro-cotovelo e quadril Joelho, permitindo a montagem precisa de poses mesmo em cenários complexos, como multidões ou oclusões.

A arquitetura do OpenPose combina mapas de confiança para localizar pontos-chave e PAFs para associá-los, refinando as estimativas em um pipeline iterativo. Esse método reduz a complexidade computacional em relação a abordagens top-down, que dependem da detecção inicial de pessoas antes da estimativa de pose. O OpenPose também suporta a detecção de pés, mãos e rosto, sendo amplamente utilizado em áreas como esportes, animações, vigilância e interação humano-computador. (Cao *et al.*, 2019)

PoseNet é um modelo de estimativa de pose humana em tempo real, implementado com TensorFlow.js, pelo fato de funcionar diretamente em navegador, foi levado em consideração para o escopo do trabalho, que visa uma aplicação leve e funcional em aparelhos móveis. Ele utiliza uma arquitetura baseada em redes convolucionais para identificar a localização de 17 pontos-chave do corpo humano, como ombros, joelhos e tornozelos, em imagens ou vídeos. Para isso, o modelo combina heatmaps que destacam regiões de alta probabilidade dos keypoints com vetores de deslocamento que ajustam a posição desses pontos com precisão.

O PoseNet oferece dois modos principais de operação: single-pose e multi-pose. No modo single-pose, o modelo detecta uma única pessoa na cena, priorizando desempenho e simplicidade para aplicações mais diretas. Já no modo multi-pose, ele utiliza algoritmos de associação baseados em grafos para identificar múltiplas pessoas em uma imagem, agrupando as partes do corpo de cada indivíduo, mesmo em situações complexas com interações ou oclusões.

A integração com o navegador e sua eficiência tornam o PoseNet ideal para uma ampla gama de aplicações, como jogos, monitoramento de exercícios, interação humano-computador e realidade aumentada. Essa combinação de simplicidade e desempenho oferece uma solução robusta para aplicações que exigem processamento em tempo real diretamente no dispositivo do usuário. (Oved *et al.*, 2018)

2.2 MÉTRICAS DE AVALIAÇÃO DE ESTIMATIVA DE POSTURA

Nesta seção, são apresentadas as métricas utilizadas para avaliar os modelos nos experimentos descritos no Capítulo 4. A aplicação dessas métricas nos estudos de caso visa fornecer subsídios para uma análise aprofundada das limitações e do desempenho das redes abordadas neste trabalho.

2.2.1 Object Keypoint Similarity

O OKS (Object Keypoint Similarity) é uma métrica que avalia a precisão no posicionamento dos keypoints em tarefas de estimativa de pose. Ele é calculado com base na distância euclidiana entre os pontos preditos pelas redes neurais e os pontos de referência anotados manualmente (ground truth), sendo essa distância normalizada pela escala da pessoa (Badiola-Bengoia; Mendez-Zorrilla, 2021).

Essa métrica oferece uma avaliação quantitativa da qualidade das previsões feitas pelo modelo, indicando o quão próximos os keypoints estimados estão em relação aos valores de referência, considerando variações de escala e posição no contexto da imagem analisada. O cálculo se dá por meio da Equação 1.

$$OKS = \exp\left(\frac{-d^2}{8\sigma^2 * A_{bb}}\right) \quad (1)$$

Onde:

- d é a distância euclidiana entre os keypoints de referência e inferidos pelo modelo, calculada pela expressão $d = \sqrt{(x_{true} - x_{pred})^2 + (y_{true} - y_{pred})^2}$
- x_{true} e y_{true} são as coordenadas cartesianas dos keypoints de referência
- x_{pred} e y_{pred} são as coordenadas cartesianas dos keypoints inferidos pela rede
- σ é o fator de escala para cada ponto-chave específico, relacionado à variabilidade esperada para aquele ponto, isto é, a tolerância de posicionamento em relação à visibilidade da junta
- A_{bb} é a área da bounding box ou caixa delimitadora do objeto

2.2.2 Erro relativo de dispersão OKS

A métrica de erro relativo de dispersão no OKS foi proposta para facilitar a visualização dos resultados obtidos, dado que os valores normalizados do OKS frequentemente se aproximam de 1, independentemente do desempenho real do modelo. Para mitigar essa limitação, adotaram-se as seguintes estratégias:

- Cálculo com σ fixo: O OKS com $\sigma = 0.1$ é utilizado como uma métrica de precisão, avaliando a capacidade do modelo de posicionar os keypoints sob um critério estrito. Um σ fixo e baixo reduz a tolerância para discrepâncias, tornando

a métrica mais rigorosa e sensível a desvios. Esse cálculo reflete o desempenho do modelo em condições adversas, destacando o alinhamento entre predições e referência.

- Erro relativo entre OKS_{var} e $OKS_{0.1}$: A diferença entre o OKS com σ variável (OKS_{var}) e o OKS com $\sigma = 0.1$ é interpretada como uma medida de erro relativo, indicando a perda de precisão devido ao relaxamento da tolerância. Um erro relativo baixo sugere que o modelo mantém alta precisão mesmo sob condições estritas, enquanto valores mais elevados evidenciam degradação no desempenho à medida que o critério se torna mais rigoroso.

Essa abordagem oferece uma medida quantitativa da sensibilidade e precisão no posicionamento dos keypoints, avaliando os modelos em diferentes níveis de tolerância. O cálculo do erro relativo fornece, assim, uma "distância da idealidade", permitindo avaliar quão próximo o modelo está de um posicionamento perfeito, representado por um OKS ideal com $\sigma = 0.1$. O cálculo da métrica se dá por meio da Equação 2.

$$Erro = \frac{|OKS_{var} - OKS_{0.1}|}{OKS_{var}} \quad (2)$$

Onde:

- OKS_{var} é o valor de OKS calculado normalmente, com σ variável por critério de visibilidade do ponto-chave
- $OKS_{0.1}$ é o valor de OKS para $\sigma = 0.1$ fixo em todos os pontos-chave

2.2.3 Mean Per Part Error

O Mean Per Part Error (MPPE) é uma métrica derivada do Percentage of Correct Parts (PCP), que avalia se um par de juntas foi detectado corretamente com base em um limiar de distância euclidiana, conforme descrito por (Josyula; Ostadabbas, 2021).

Embora o PCP forneça uma medida percentual da detecção correta de pares, essa informação isolada pode ser pouco expressiva. Por outro lado, o MPPE calcula o erro relativo entre os pares de juntas, permitindo avaliar a consistência do modelo na estimativa do tamanho dos membros em relação ao esqueleto de referência. Essa métrica indica o desvio proporcional na estimativa do modelo, determinando se os elos mantêm dimensões coerentes com o ground truth.

No contexto deste estudo, o MPPE é calculado com base no erro relativo de posicionamento entre pares de juntas que formam as partes do corpo. Essa abordagem representa a capacidade do modelo em preservar a coesão estrutural do esqueleto, fornecendo uma avaliação mais detalhada do desempenho do modelo em manter as proporções anatômicas corretas. O cálculo se dá por meio da Equação 3

$$Erro = \frac{|d_{gt} - d_{pred}|}{d_{gt}} \quad (3)$$

Onde:

- d_{gt} é a distância eclidiana entre 2 pontos-chave adjacentes da referência, calculada por meio da expressão $d_{gt}(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$
- d_{pred} é a distância eclidiana entre 2 pontos-chave adjacentes inferidos pelo modelo, calculada por meio da expressão $d_{pred}(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$
- i e j são índices de coordenadas cartesianas x e y dos keypoints adjacentes

2.3 TRABALHOS CORRELATOS

Dentre os trabalhos relacionados ao tema, destaca-se o estudo de Sugawara (2022), que abordou a avaliação de modelos como MediaPipe e MoveNet para aplicações móveis, focando na análise tridimensional de exercícios físicos. A proposta de Sugawara parte da premissa de que a execução inadequada de atividades físicas pode causar lesões musculares, ósseas e ligamentares. Assim, o trabalho propõe um sistema capaz de avaliar geometricamente exercícios a partir de múltiplas perspectivas, utilizando uma rede neural recorrente para classificar exercícios como "agachamento" e "afundo" e implementar uma dupla verificação tridimensional. Essa abordagem busca suprir lacunas dos modelos existentes, que frequentemente limitam-se a análises bidimensionais de uma única perspectiva, como frontal ou lateral, inadequadas para exercícios complexos.

Por outro lado, como destacado em artigos de revisão, como o de Badiola-Bengoia e Mendez-Zorrilla (2021) e Dubey e Dixit (2023), a aplicabilidade da estimativa de pose humana (HPE) no esporte ainda está em seus estágios iniciais, com inúmeras possibilidades de pesquisa. O número de esportes e aplicações em que HPE não foi aplicada, ou foi pouco explorada, é vasto. Além disso, conforme discutido, o desenvolvimento de sistemas que priorizem aspectos como rapidez, leveza, setups especializados para problemas concretos e o uso de hardware de baixo custo representa uma oportunidade significativa de estudo. Entre os desafios enfrentados por diversos autores na aplicação de HPE em movimentos específicos ou esportes, destaca-se o problema das oclusões, o que indica um campo promissor para pesquisas voltadas à mitigação desses efeitos, sem se limitar apenas ao aumento da precisão na predição das posições das articulações.

Em contraste, este trabalho foca na sistematização de um procedimento de avaliação de modelos pré-treinados de estimativa de pose humana, sem realizar treinamento adicional ou desenvolvimento de novas redes neurais. Enquanto Sugawara propõe a criação de um novo sistema para análise tridimensional e classificação de exercícios, a presente pesquisa objetiva avaliar o desempenho de modelos existentes

em seu estado padrão de distribuição. A metodologia proposta aqui busca identificar as limitações e o potencial desses modelos em cenários de capacidade computacional restrita, como dispositivos móveis, validando métricas específicas através de estudos de caso.

Dessa forma, as abordagens se complementam. Enquanto Sugawara explora o desenvolvimento de novas ferramentas para aprimorar a análise geométrica de exercícios, e artigos de revisão destacam os desafios e as oportunidades ainda inexploradas da HPE no esporte, este trabalho contribui para a avaliação crítica de tecnologias existentes, fornecendo subsídios para sua aplicação prática em diferentes contextos, além de levantar questões sobre a adequação dos modelos em cenários reais.

3 METODOLOGIA

Este capítulo apresenta a metodologia adotada no desenvolvimento do presente trabalho, detalhando os procedimentos, ferramentas e experimentos realizados. A Figura 6 ilustra, por meio de um fluxograma, as etapas do processo de pesquisa e desenvolvimento conduzidas para estruturar o comparativo entre redes neurais de estimativa de pose.

Inicialmente, foi realizada uma pesquisa sobre redes neurais voltadas à estimativa de postura humana, resultando na seleção de abordagens amplamente reconhecidas por sua popularidade e desempenho. Os modelos escolhidos foram:

- YOLO, estável na versão 8 durante a realização deste trabalho;
- MediaPipe, cuja aplicação descrita por Dawn (2022) serviu como referência inicial para o projeto;
- OpenPose, modelo desenvolvido por pesquisadores da Carnegie Mellon School of Computer Science;
- MoveNet e PoseNet, ambos pertencentes à biblioteca TensorFlow.

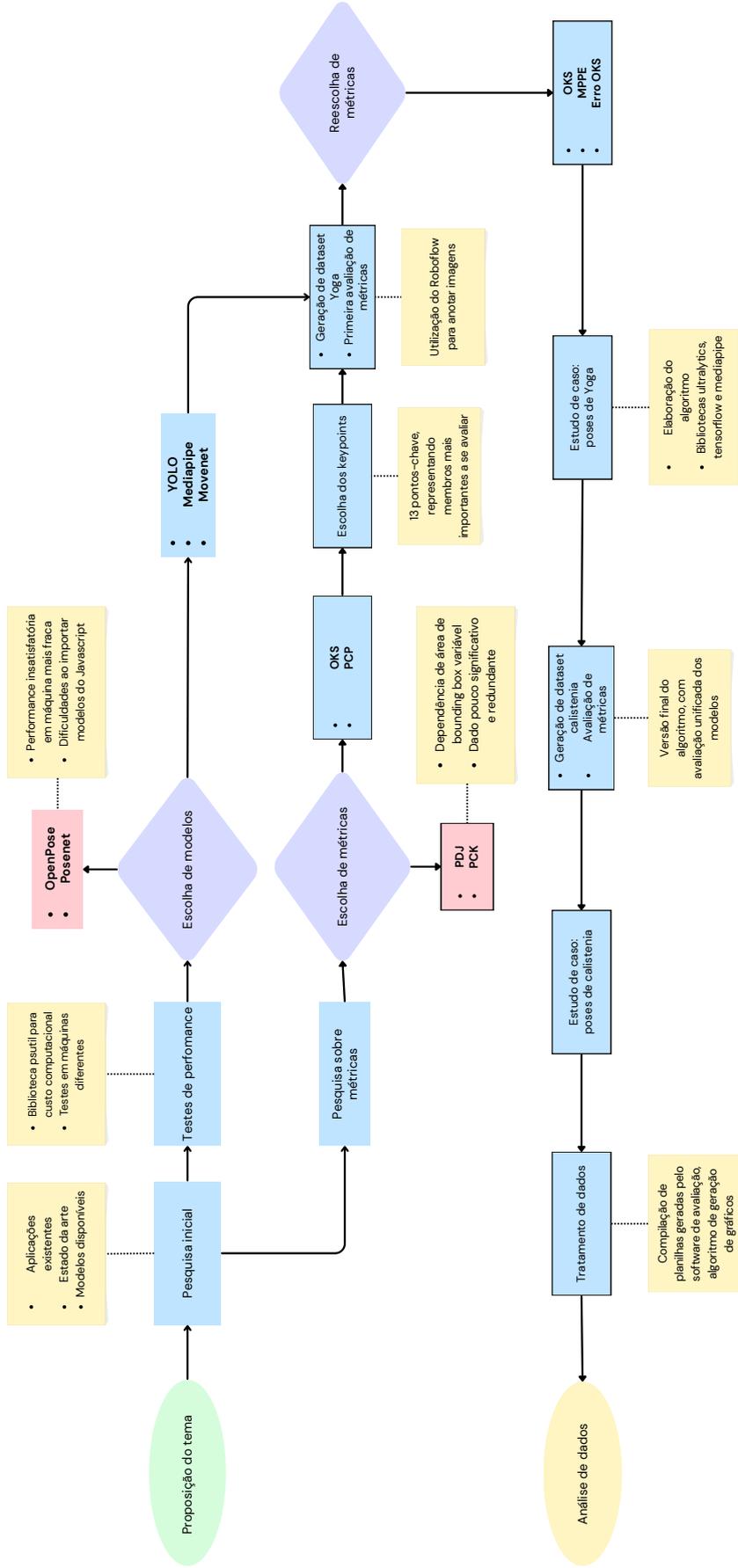
Durante os testes preliminares, constatou-se que o OpenPose apresentava desempenho inadequado em máquinas de menor capacidade computacional. Testado em um computador com processador AMD 5600X e placa de vídeo RTX 3070, o modelo atingiu 30 a 40 quadros por segundo ao inferir diretamente sobre vídeos capturados pela câmera. No entanto, ao ser executado em um laptop com processador Intel 7300HQ e placa de vídeo GTX 1050Ti, o desempenho caiu para apenas 7 quadros por segundo, tornando inviável sua aplicação em tempo real para prevenção de lesões.

Adicionalmente, a implementação do PoseNet foi descartada devido à complexidade de sua adaptação do ambiente JavaScript para Python. Assim, o MoveNet foi escolhido por sua integração nativa com Python, facilitando a experimentação.

Em seguida, foi conduzido um teste para avaliar a eficiência computacional dos modelos utilizando a biblioteca **psutil** em Python. O experimento, cujo algoritmo está disponível no repositório (Nina, 2024b) do Github, consistiu em realizar inferência sobre um vídeo de aproximadamente cinco minutos de duração, contendo uma pessoa executando flexões de braço. Os resultados médios, relacionados ao uso de processador e memória, estão apresentados no Quadro 1.

Para realizar o estudo comparativo, foram necessários dados anotados que serviriam como referência (ground truth) para o cálculo de métricas. Os pontos-chave foram definidos considerando as articulações principais de exercícios de corpo inteiro,

Figura 6 – Fluxograma do desenvolvimento do trabalho.



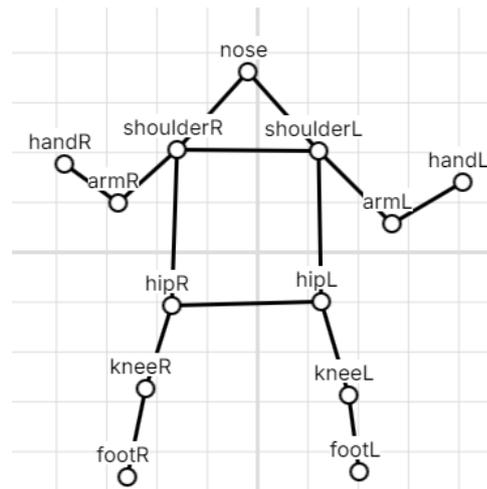
Fonte: Autor (2024).

Quadro 1 – Resultados de custo computacional sobre inferência em vídeo.

Recurso	YOLO	Mediapipe	Movenet
Utilização de processador	16.4%	10.5%	3.73%
Memória	7438.45 MB	86.76 MB	6523.57 MB

Fonte: Autor(2024).

Figura 7 – Diagrama de pontos-chave seleccionados no Roboflow.



Fonte: Autor (2024).

Figura 8 – Anotação de keypoints sobre imagem no Roboflow.



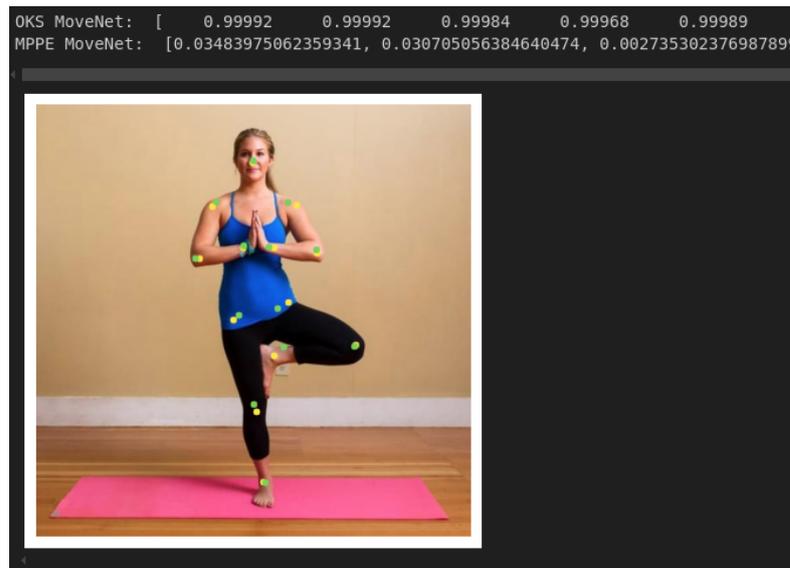
Fonte: Autor (2024).

totalizando 13 keypoints: o nariz (representando o centro da cabeça) e 12 articulações principais do corpo humano, conforme ilustrado na Figura 7.

A anotação dos pontos-chave foi realizada manualmente utilizando a interface do software Roboflow, que permitiu gerar um dataset compatível com os algoritmos de cálculo de métricas. Um exemplo dessa anotação pode ser observado na Figura 8.

Para o primeiro experimento, foi utilizado um dataset de 16 imagens de poses de Yoga, obtidas na plataforma Roboflow. As poses de Yoga representam um desafio comum para pose estimation, com poses geralmente bidimensionais e vistas de frente, porém com algumas flexões agudas de membros. As imagens foram redimen-

Figura 9 – Exemplo de saída do código.



Fonte: Autor (2024).

cionadas para 640x640 pixels. Para garantir uniformidade no cálculo de métricas, foi adotado um valor aproximado de 53% da área da imagem como estimativa da área segmentada do objeto, similar ao método descrito por Dutta e Dawn (2023), que sugere a aplicação de um fator heurístico de escala (0,53) sobre a área da bounding box. No entanto, diferentemente do caso da fonte mencionada, em que a área da bounding box é utilizada, optou-se pelo uso da área da imagem, visto que as informações sobre bounding boxes não estavam disponíveis neste caso. Um algoritmo unificado (Nina, 2024a) de avaliação de métricas para os três modelos selecionados foi desenvolvido; nele, os keypoints das labels do dataset foram extraídos e utilizados como base para comparação com os resultados inferidos pelos modelos testados.

Os modelos realizaram inferência sobre as imagens, e os keypoints previstos foram extraídos e comparados aos valores de referência. Como resultado, foi gerada uma saída com um comparativo visual entre os pontos inferidos e os ground truth, conforme exemplificado na Figura 9, além do armazenamento das métricas em planilhas para posterior análise estatística, incluindo identificação de outliers e articulações mais suscetíveis a erro.

O segundo experimento utilizou um dataset autoral, composto por imagens de três exercícios específicos de calistenia, isto é, atividades de treino de resistência utilizando apenas o peso corporal, na forma dos movimentos de agachamento, flexão de braço e avanço, capturados em três ângulos distintos, visando analisar o comportamento das redes em diferentes situações pouco ideais como na captura das poses por meio de celular, havendo muitas oclusões. O dataset foi formado por nove imagens, também anotadas manualmente no Roboflow.

Os resultados das inferências realizadas pelos modelos foram comparados às anotações de referência, considerando as métricas previamente definidas. Esse experimento visou avaliar o desempenho dos modelos em cenários mais controlados e com poses diretamente relacionadas a exercícios físicos.

4 EXPERIMENTOS E RESULTADOS

Neste capítulo, são apresentados os experimentos realizados e seus respectivos resultados. A análise inclui gráficos do tipo box plot, que ilustram a dispersão dos erros em termos de OKS e MPPE. Esses gráficos permitem uma visão clara sobre as deficiências na detecção de juntas, com destaque para os valores atípicos (outliers), que serão explorados em maior profundidade devido à sua relevância na avaliação do desempenho dos modelos. Nas discussões sobre os casos mais críticos, os pontos representados em amarelo correspondem aos keypoints de referência, anotados manualmente, enquanto os pontos em verde representam aqueles inferidos pelos modelos analisados.

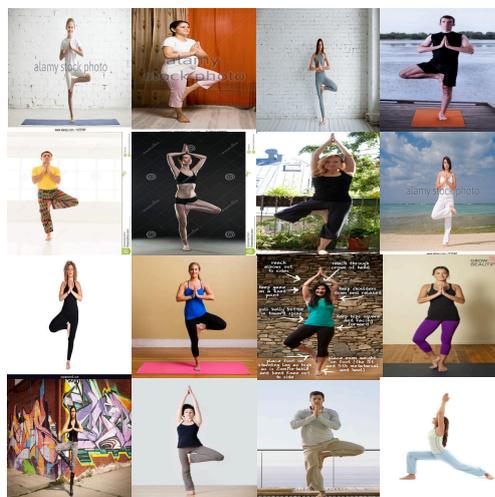
4.1 ESTUDO DE CASO 1

Foi realizado um primeiro estudo de caso utilizando as 16 imagens da Figura 10 contendo poses estáticas de Yoga. Essas poses foram selecionadas por representarem desafios comuns para os modelos de estimativa de postura, como articulações parcialmente ocultas, ângulos extremos e posições assimétricas.

4.1.1 YOLO

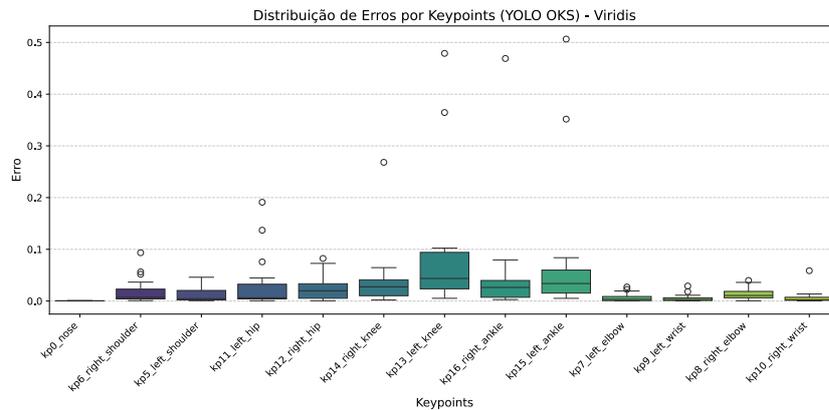
Na Tabela 2 do Apêndice A, observa-se que os resultados médios de OKS_{var} são predominantemente superiores a 0,99. Essa característica dificulta a identificação

Figura 10 – Dataset com imagens de Yoga utilizado.



Fonte: Autor (2024).

Figura 11 – Box plot do $Erro_{OKS}$ para o YOLO no estudo de caso 1.



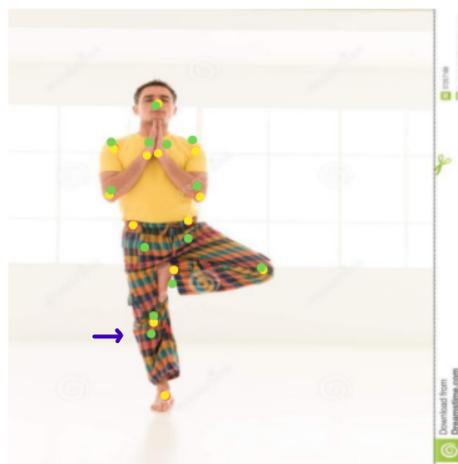
Fonte: Autor (2024).

de limitações do modelo, mesmo ao recalcular a métrica utilizando um σ reduzido, conforme observado na Tabela 3 do Apêndice B, onde as médias permanecem elevadas, restringindo a análise detalhada dos dados.

Para melhorar a escala de análise do erro, optou-se pelo cálculo do erro relativo em relação ao OKS. A Figura 11 apresenta um box plot desses resultados, permitindo uma visualização mais clara da dispersão dos erros.

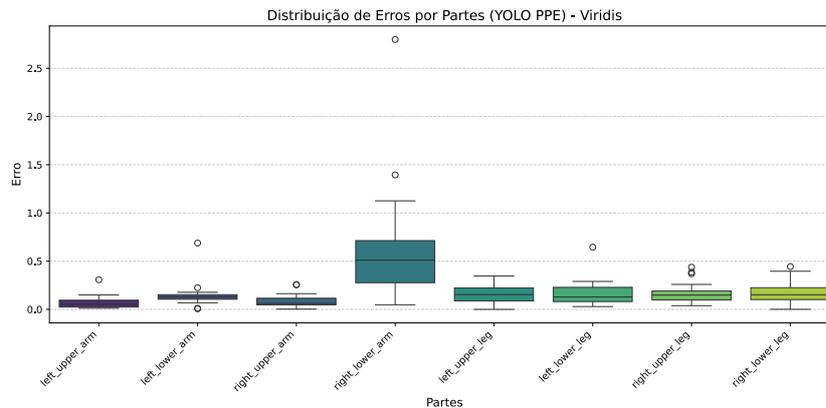
Durante os experimentos foi identificada uma dificuldade do modelo em localizar articulações inferiores, como joelhos e calcanhares. Imagens contendo a postura da árvore, caracterizada por uma flexão acentuada do joelho, apresentaram os piores desempenhos. Curiosamente, o maior outlier exibido no gráfico da Figura 11 não ocorreu devido à flexão, mas à perna estendida, sugerindo que o pré-treinamento foi realizado em datasets predominantemente compostos por posturas simétricas.

Figura 12 – Perda de ponto-chave no calcanhar esquerdo pelo YOLO.



Fonte: Autor (2024).

Figura 13 – Box plot do MPPE para o YOLO no estudo de caso 1.



Fonte: Autor (2024).

Embora visualmente seja evidente a perda de precisão em algumas inferências, o cálculo do OKS em sua configuração padrão é menos sensível a erros localizados. Para uma análise mais detalhada, recalculou-se a métrica com uma tolerância reduzida ($\sigma = 0.1$), destacando perdas significativas. No caso da Figura 12, por exemplo, $OKS_{var} = 0.9813$, indicando um desempenho aparentemente alto. Contudo, com o σ ajustado, o valor reduziu-se para 0.4843, representando uma queda de 50,65%, evidenciando a fragilidade do modelo em cenários mais restritivos.

O MPPE apresentou comportamento inverso ao erro do OKS, como mostra a Figura 13, com maior acentuação em membros superiores, especialmente nos antebraços. Observa-se um aumento nas médias de erro, indicando a maior sensibilidade dessa métrica em identificar desalinhamentos nos membros, conforme ilustrado na Figura 14.

Na Figura 14, é possível observar um acúmulo de erros elevados no braço direito, que está oculto pelo esquerdo e apresenta desalinhamento nos keypoints de punho e cotovelo. Essa discrepância é amplificada pela natureza da métrica MPPE, destacando mais intensamente inconsistências nos alinhamentos e localizações dos pontos-chave. Esse resultado reforça o potencial do MPPE em detectar falhas que o OKS pode não evidenciar de forma tão expressiva. O outlier identificado com $MPPE = 2.7984$ indica uma estimativa incorreta do modelo, refletindo uma má localização dos pontos-chave no braço direito.

4.1.2 Mediapipe

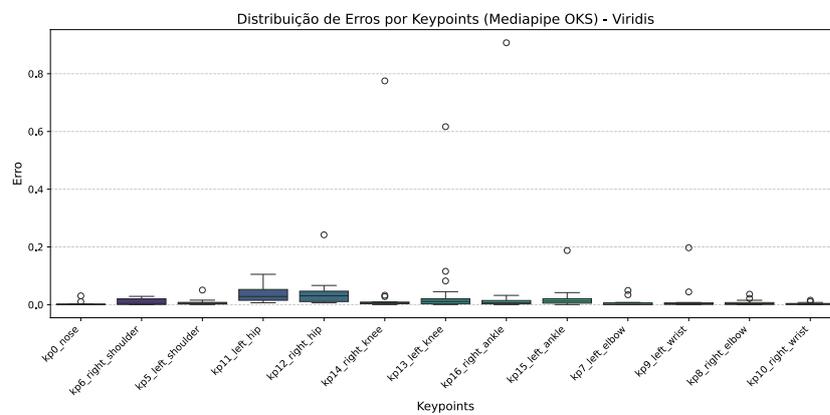
Comportamentos semelhantes foram observados no Mediapipe, evidenciando maior concentração de erros nos outliers, embora as médias permaneçam elevadas e próximas às do YOLO. A Figura 15 apresenta o box plot do OKS para este modelo, revelando limitações na detecção de poses com articulações em posições extremas.

Figura 14 – Pior caso do MPPE para o YOLO no estudo de caso 1



Fonte: Autor (2024).

Figura 15 – Box plot do $Erro_{OKS}$ para o Mediapipe no estudo de caso 1.



Fonte: Autor (2024).

Figura 16 – Perda de ponto-chave no calcanhar direito pelo Mediapipe.



Fonte: Autor (2024).

A Figura 16, por sua vez, ilustra um exemplo específico com a postura da árvore, espelhada e com um deslocamento mais acentuado no quadril. Neste exemplo, destaca-se um erro significativo na representação gráfica dos pontos inferidos, que afeta principalmente a perna flexionada. A rede apresenta dificuldade em localizar corretamente o joelho, confundindo-o com o keypoint do pé direito. Essa imprecisão compromete a integridade da estimativa para poses mais desafiadoras, especialmente aquelas com flexões agudas ou posições assimétricas.

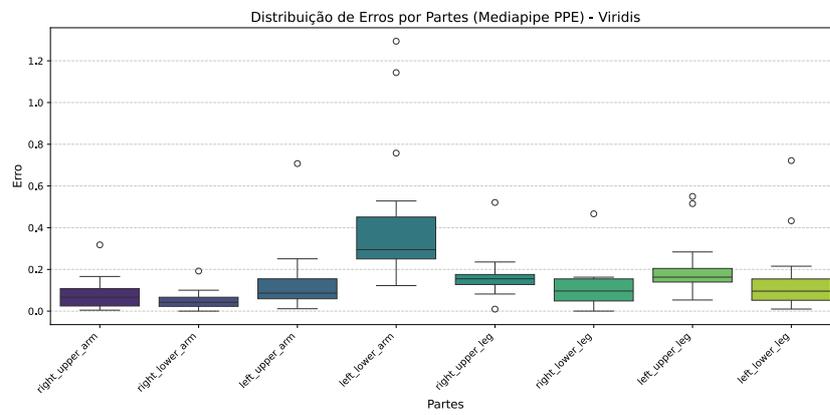
Para a Figura 16, o OKS atribui ao pé direito um score de 0.9384, valor que pode ser interpretado como satisfatório caso a métrica seja tratada de forma similar à acurácia. Entretanto, ao reduzir a tolerância para uma análise mais rigorosa, a pontuação cai drasticamente para 0.0869, indicando uma perda de 90,75%. Essa discrepância evidencia que, embora o modelo seja capaz de estimar poses com boa precisão em situações gerais, ele apresenta fragilidades significativas quando avaliado em cenários de maior exigência ou com ajustes na tolerância.

e acordo com o gráfico da Figura 17, observa-se que a métrica MPPE identificou uma quantidade significativamente maior de casos críticos no Mediapipe em comparação ao YOLO e Movenet, que será discutido posteriormente. Notam-se outliers mais elevados nos keypoints relacionados ao braço e antebraço esquerdos, bem como na coxa e canela esquerdas.

Embora o caso de erro máximo apresente semelhanças com o ilustrado na Figura 14, a análise será focada no caso apresentado na Figura 18, que se destaca por apresentar outliers em mais da metade de suas medições, abrangendo a totalidade da perna e do braço esquerdos.

Na Figura 18, é evidente o desalinhamento de juntas e elos na metade direita da imagem (considerando o referencial espelhado). Há keypoints localizados fora do

Figura 17 – Box plot MPPE para o Mediapipe no estudo de caso 1.



Fonte: Autor (2024).

Figura 18 – Pior caso apresentado pelo MPPE no Mediapipe.



Fonte: Autor (2024).

corpo, incluindo uma translação significativa do ponto correspondente ao joelho, posicionado próximo ao calcanhar. Os resultados de MPPE para este caso estão na Tabela 1.

Tabela 1 – Valores de MPPE para o outlier da Figura 18.

Parte	MPPE
<i>left_upper_arm</i>	0,7073
<i>left_lower_arm</i>	1,2939
<i>left_upper_leg</i>	0,5154
<i>left_lower_leg</i>	0,7216

Fonte: Autor (2024).

Essa pose é particularmente distinta das demais, uma vez que os braços estão de encontro um ao outro à frente do peito, enquanto o ângulo da câmera é parcialmente lateralizado. Além disso, a presença de uma marca d'água no centro da imagem pode ter influenciado negativamente o desempenho do Mediapipe, dificultando a detecção precisa das juntas. Esses fatores combinados podem explicar o elevado número de outliers registrados nessa situação.

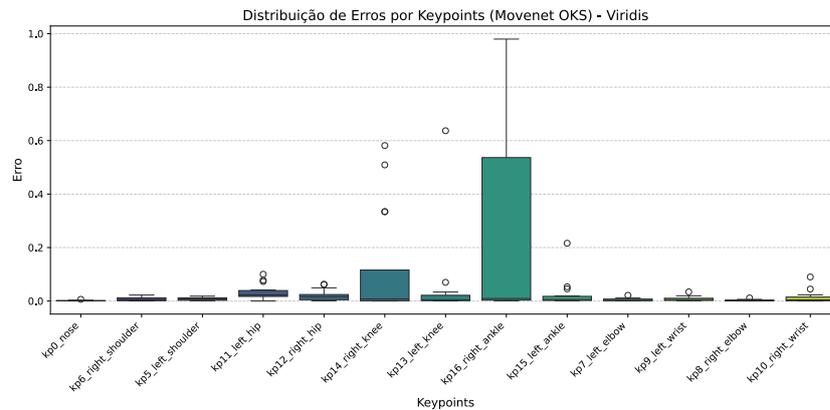
4.1.3 Movenet

O Movenet apresentou padrões semelhantes aos modelos anteriores, com concentrações de erro mais acentuadas nas articulações inferiores, especialmente no calcanhar direito. Essa articulação destacou-se como um outlier em relação às demais redes, exibindo uma média de erro aproximadamente quatro vezes maior do que as observadas no YOLO e no Mediapipe. A Figura 19 apresenta o box plot dos valores de OKS para o Movenet, evidenciando suas limitações na detecção de poses desafiadoras.

A análise revelou que os erros são mais pronunciados em imagens que apresentam flexões ou assimetrias acentuadas. O caso ilustrado na Figura 20, por exemplo, refere-se à mesma imagem-problema identificada no Mediapipe. O OKS calculado para o calcanhar direito foi de 0.9009, mas, ao adotar um valor de σ menor (0.1), o score caiu para 0.0181, representando um erro relativo de 98%.

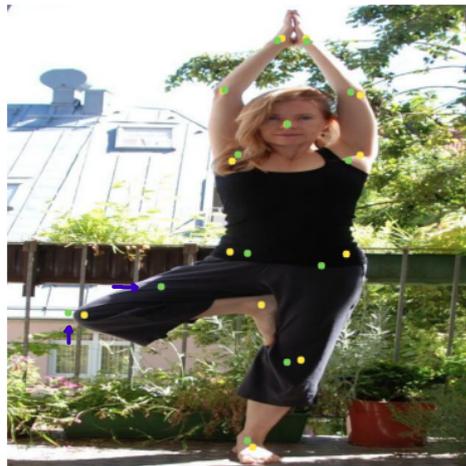
A acentuada redução no desempenho indica que o Movenet pode ter sido treinado principalmente com imagens de posturas simétricas, limitando sua capacidade de generalizar para poses complexas ou atípicas. Essa limitação ressalta a necessidade de um treinamento adicional, direcionado a cenários mais diversos, para adaptar a ferramenta a diferentes tipos de atividades físicas e melhorar sua precisão na detecção de articulações inferiores em situações desafiadoras.

Figura 19 – Box plot do $Erro_{OKS}$ para o Movenet no estudo de caso 1.



Fonte: Autor (2024).

Figura 20 – Perda de ponto-chave no calcanhar direito pelo Movenet.



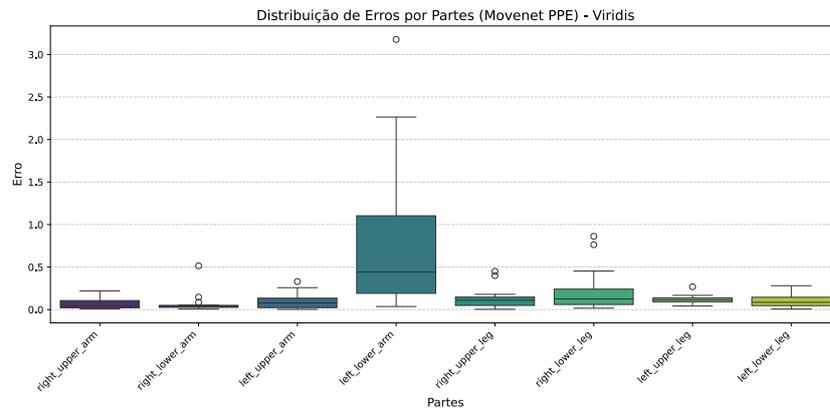
Fonte: Autor (2024).

Na Figura 21, observa-se o box plot dos erros MPPE para o modelo Movenet, destacando-se o antebraço esquerdo, que apresenta uma média consideravelmente superior em comparação aos demais elos, além do outlier mais elevado registrado. Este caso específico está ilustrado na Figura 22.

Na Figura 22, dois fenômenos chamam atenção: os erros nos keypoints dos braços e uma estimativa equivocada do Movenet, que interpretou as pernas como estando alinhadas. Apesar do erro evidente na coxa direita, o modelo apresentou um score melhor, de somente 0,2687 de erro nesse membro quando comparado ao braço esquerdo, cujo resultado foi de $MPPE = 3,1769$.

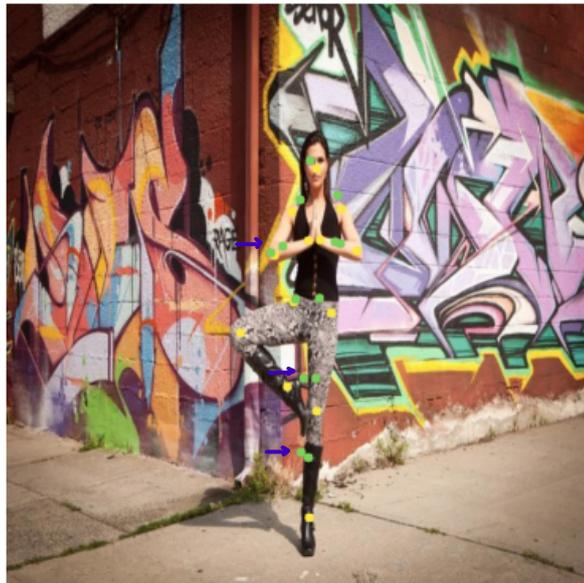
Esse resultado pode ser explicado pela forma como a métrica MPPE avalia os erros. A métrica considera o tamanho relativo do elo inferido em relação ao tamanho esperado, e não apenas a distância de posicionamento entre os keypoints inferidos e a ground truth. No caso do antebraço esquerdo, a inferência do modelo resultou em

Figura 21 – Box plot do MPPE para o Movenet no estudo de caso 1.



Fonte: Autor (2024).

Figura 22 – Erro MPPE destacado para o modelo Movenet em pose de yoga.



Fonte: Autor (2024).

uma sobreposição quase completa dos keypoints, reduzindo drasticamente o tamanho do elo. Esse erro, proporcionalmente, foi maior do que o da perna direita, que, embora deslocada, manteve um tamanho similar ao esperado.

Além disso, na perna direita, o modelo a interpretou como estendida, apesar de estar dobrada na realidade. Isso sugere que o Movenet apresentou dificuldades em capturar adequadamente a postura complexa da figura, especialmente em regiões de sobreposição ou ângulos não usuais.

Figura 23 – Dataset com frames de exercícios de calistenia.



Fonte: Autor (2024).

4.2 ESTUDO DE CASO 2

Para avaliar o desempenho das redes neurais em uma aplicação voltada a exercícios de calistenia realizados em condições distintas e desafiadoras, como aquelas associadas ao uso de smartphones, foram selecionadas 9 imagens representativas, ilustradas na Figura 23. Essas imagens capturam momentos de maior flexão dos músculos-alvo de cada exercício durante sua execução.

As fotos foram obtidas de três ângulos distintos: frontal, lateral e traseiro. Essa abordagem visa ampliar a diversidade dos esqueletos anotados e analisar a performance dos modelos ao lidar com articulações ocultas ou parcialmente visíveis, como quadris e ombros.

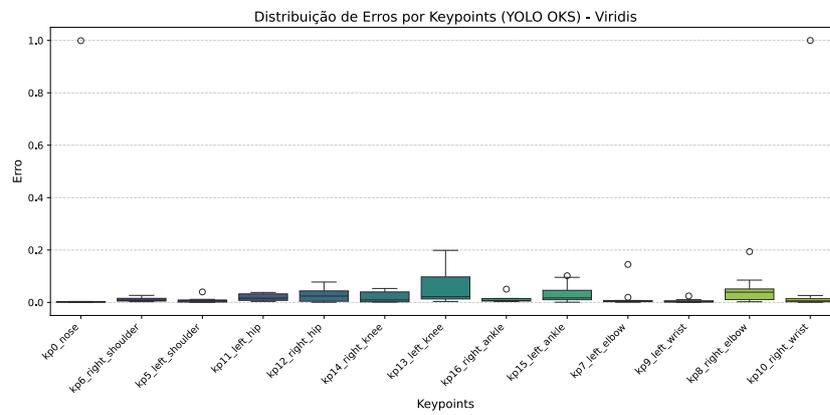
O estudo tem como objetivo identificar limitações dos modelos em cenários que envolvem posicionamentos adversos da câmera, destacando dificuldades na detecção de juntas sobrepostas ou obstruídas.

4.2.1 YOLO

Diferentemente do estudo anterior, o acúmulo de erro em joelhos e calcanhares foi reduzido. Contudo, duas articulações que antes não apresentavam desvios significativos destacaram-se como outliers na Figura 24: nariz e punho direito, ambos com pontuações de OKS muito inferiores ao padrão, geralmente acima de 0,9.

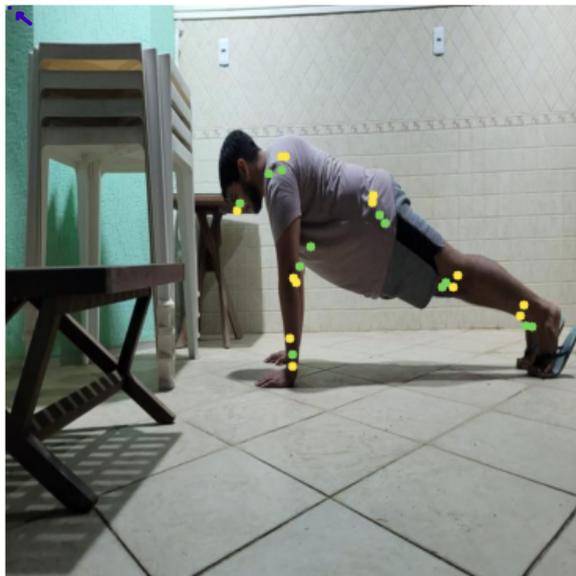
Esse comportamento decorre de um novo padrão observado: keypoints completamente ocultos não são identificados pelo modelo. No caso do punho direito, a articulação sobreposta por outro ponto não foi localizada. Com isso, os valores de OKS aproximaram-se de zero para σ fixo.

Figura 24 – Box plot do $Erro_{OKS}$ para o YOLO no estudo de caso 2.



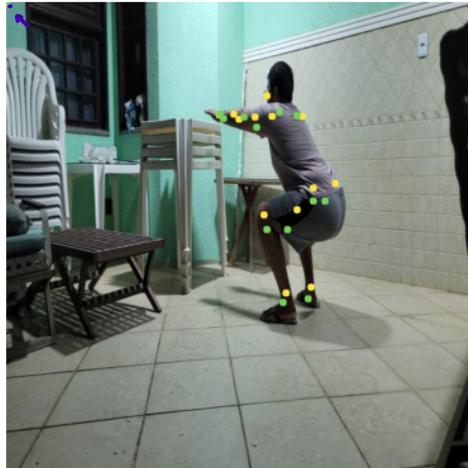
Fonte: Autor (2024).

Figura 25 – Ponto-chave da mão direita é perdido pelo YOLO.



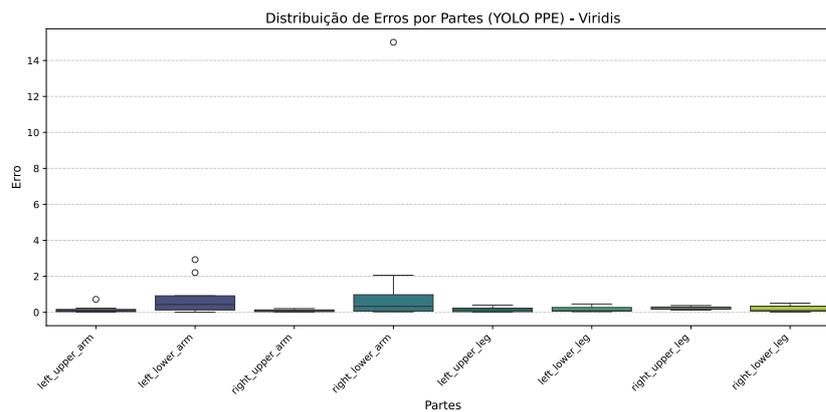
Fonte: Autor (2024).

Figura 26 – Ponto chave do nariz é perdido pelo YOLO.



Fonte: Autor (2024).

Figura 27 – Box plot MPPE para o YOLO no estudo de caso 2.



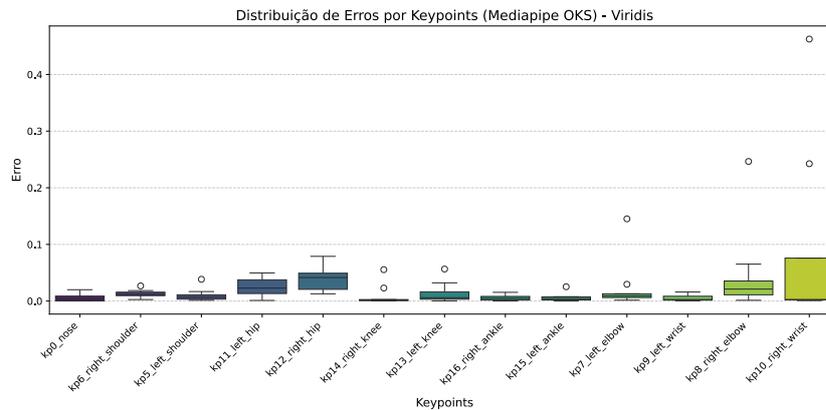
Fonte: Autor (2024).

Conforme ilustrado nas Figuras 25 e 26, keypoints não detectados são posicionados, por padrão, no canto superior esquerdo da imagem. Nos casos analisados, os valores de OKS_{var} foram 0,2984 e 0,7026, respectivamente. Entretanto, ao utilizar σ fixo, os resultados tornaram-se insignificantes, abaixo de 10^{-3} . Assim, o erro relativo para o punho foi $Erro_{OKS} = 0,9999981811$, enquanto para o nariz, $Erro_{OKS} = 0,9990554756$.

Conforme apresentado no gráfico da Figura 27, identifica-se um outlier com valor significativamente acima do padrão. Este caso corresponde à situação ilustrada na Figura 25, em que o modelo falhou em detectar o keypoint correspondente, resultando em um cálculo incorreto do tamanho do elo.

Na ausência do keypoint correto, o modelo posicionou o ponto estimado no topo da imagem, o que gerou um tamanho de elo muito superior ao esperado e, conseqüentemente, um erro extremamente elevado de $MPPE = 15,01$. Este comporta-

Figura 28 – Box plot do $Erro_{OKS}$ para o Mediapipe no estudo de caso 2.



Fonte: Autor (2024).

mento evidencia novamente a limitação do modelo em lidar com situações de perda de keypoints, afetando significativamente a métrica.

4.2.2 Mediapipe

Com base na Figura 28, o Mediapipe demonstrou limitações ao lidar com articulações não visíveis, como exemplificado na Figura 29, que apresenta uma vista traseira do avanço. Nessa situação, a posição exata do cotovelo e do punho direitos não pode ser determinada com precisão. Apesar disso, o modelo obteve um $OKS_{var} = 0,9835$ para o punho direito, indicando um desempenho superior em relação ao modelo anterior. No entanto, ao aplicar uma menor tolerância, o valor caiu para $OKS_{0.1} = 0,5286$.

Na Figura 29, observa-se que o modelo estimou erroneamente o antebraço direito como estendido para baixo e alinhado ao quadril, enquanto, na realidade, ele estava flexionado em direção ao punho esquerdo à frente do peito. Esse erro resultou em um valor relativo de $Erro_{OKS} = 0,4625$.

Conforme o box plot apresentado na Figura 30, é possível identificar dois outliers significativos no antebraço direito, associados às Figuras 29 e 31. Em ambos os casos, as juntas estavam ocultas, resultando em estimativas incorretas por parte do modelo.

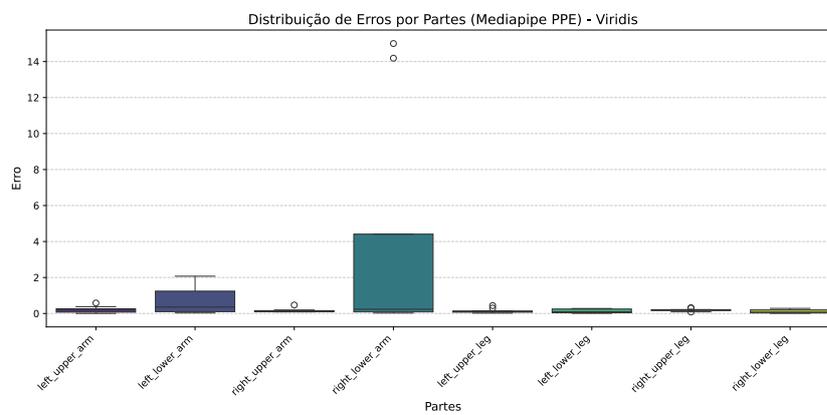
Na Figura 29 o erro, cujo resultado foi $MPPE = 15,0027$, foi causado pela suposição de que o braço estava completamente estendido, enquanto, na realidade, ele estava flexionado. Já na Figura 31, o erro decorreu do posicionamento inadequado dos keypoints do antebraço, que foram colocados muito próximos ao cotovelo, acarretando um valor de $MPPE = 14,1880$. Essa proximidade gerou uma estimativa quase nula para o tamanho do elo, incompatível com o esperado, resultando em um acúmulo significativo de erro na métrica. Esses casos evidenciam limitações do mo-

Figura 29 – Posicionamento incorreto de keypoints no braço direito, que está oculto.



Fonte: Autor (2024).

Figura 30 – Box plot do MPPE para o Mediapipe no estudo de caso 2.



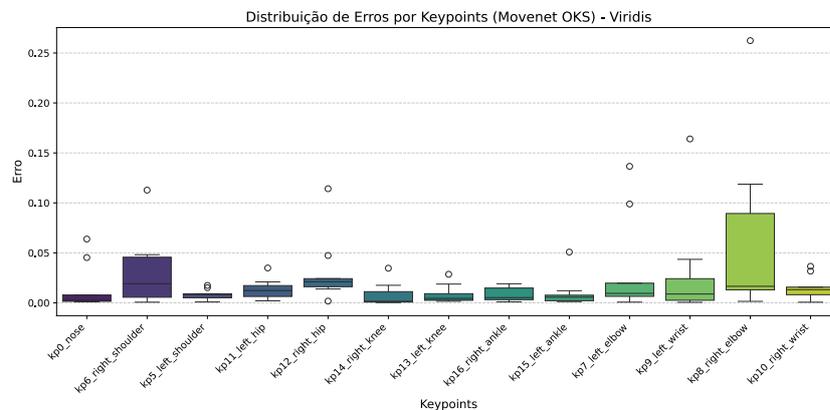
Fonte: Autor (2024).

Figura 31 – Caso de outlier do MPPE no Mediapipe.



Fonte: Autor (2024).

Figura 32 – Box plot do $Erro_{OKS}$ para o Movenet no estudo de caso 2.



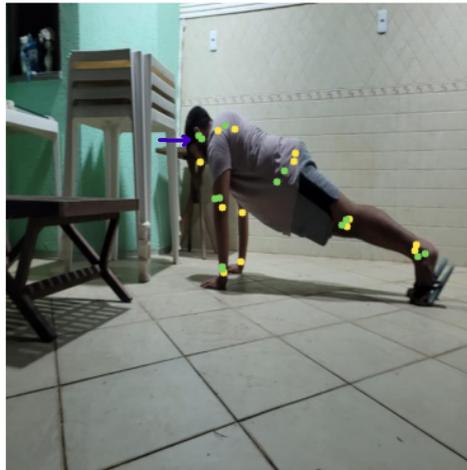
Fonte: Autor (2024).

delo em lidar com poses em que as articulações estão ocultas ou em posições não convencionais.

4.2.3 Movenet

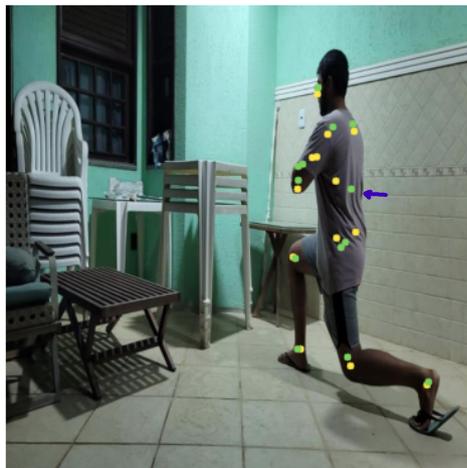
Analisando o gráfico da Figura 32, o Movenet demonstrou um erro mais acentuado no cotovelo direito, destacando-se como o ponto com maior inconsistência no experimento. Em condições normais, o menor valor de erro registrado foi para o nariz durante a execução do apoio, com $OKS_{var} = 0,9886$. Entretanto, mesmo ao reduzir a tolerância (σ) para 0,1, o erro associado a essa articulação permaneceu abaixo de 0,002, indicando impacto mínimo na análise.

Figura 33 – Posicionamento incorreto do nariz durante inferência do Movenet.



Fonte: Autor (2024).

Figura 34 – Posicionamento incorreto do cotovelo direito durante inferência do Movenet.



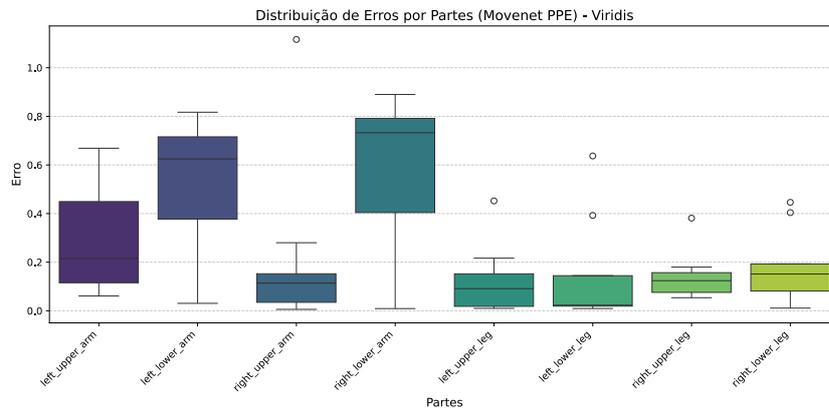
Fonte: Autor (2024).

A Figura 33 ilustra o posicionamento incorreto do ponto referente ao nariz. Apesar da discrepância visível, a variação na métrica OKS foi limitada pela baixa dispersão adotada na análise (σ de 0,26 para 0,1), o que resultou em uma diferença pouco significativa.

Por outro lado, a pontuação do cotovelo direito apresentou maior impacto negativo, particularmente na postura de avanço. Devido à visibilidade limitada da articulação, o valor do OKS caiu de $OKS_{var} = 0,9940$ para $OKS_{0.1} = 0,7332$, evidenciando a dificuldade do modelo em lidar com articulações parcialmente ocultas.

O caso mais crítico foi observado na Figura 34, com um erro relativo calculado de $Erro_{OKS} = 0,2624$. Esse resultado reforça a limitação do Movenet em generalizar corretamente a localização de juntas em situações de baixa visibilidade ou posiciona-

Figura 35 – Box plot MPPE para o Movenet no estudo de caso 2.



Fonte: Autor (2024).

mentos complexos. Um treinamento mais robusto, utilizando dados que representem articulações parcialmente ocultas, poderia melhorar o desempenho do modelo nessas condições.

A Figura 35 apresenta o gráfico de MPPE para o modelo MoveNet. Embora tenha exibido variabilidade notável, se comparado aos modelos YOLO e MediaPipe, os erros ocorreram em uma escala reduzida. O maior outlier identificado apresentou um valor de $MPPE = 1,1163$, associado ao braço direito, conforme ilustrado na Figura 34.

A origem do erro foi uma estimativa incorreta realizada pelo MoveNet, que posicionou o cotovelo direito abaixo da localização esperada. Essa discrepância resultou em um aumento no comprimento do segmento correspondente, o que elevou o valor da métrica MPPE para o referido ponto.

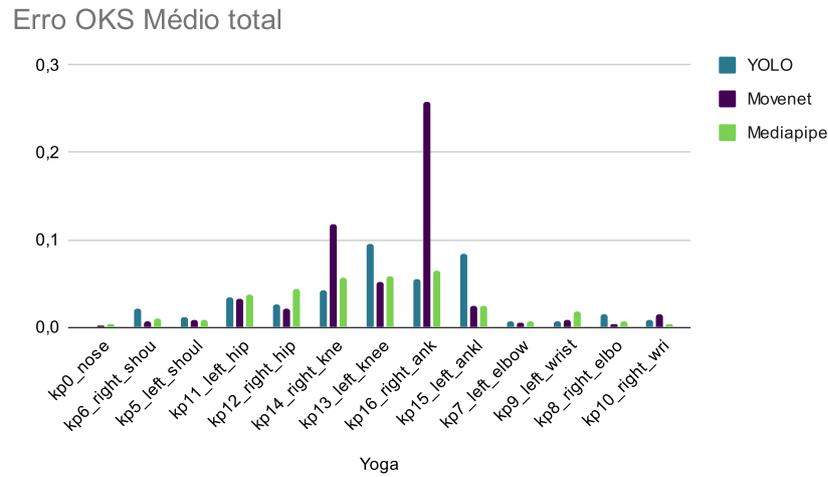
Por fim, é possível fazer um levantamento resumido dos casos mais críticos apresentados ao longo das duas últimas seções.

Quadro 2 – Maiores outliers apresentados em cada modelo para cada estudo de caso.

Modelo	Caso	$Erro_{OKS}$	MPPE
YOLO	Yoga	0,5065	2,7984
YOLO	Calistenia	0,9999	15,0100
Mediapipe	Yoga	0,9075	1,2939
Mediapipe	Calistenia	0,4625	15,0027
Movenet	Yoga	0,9800	3,1769
Movenet	Calistenia	0,2624	1,1163

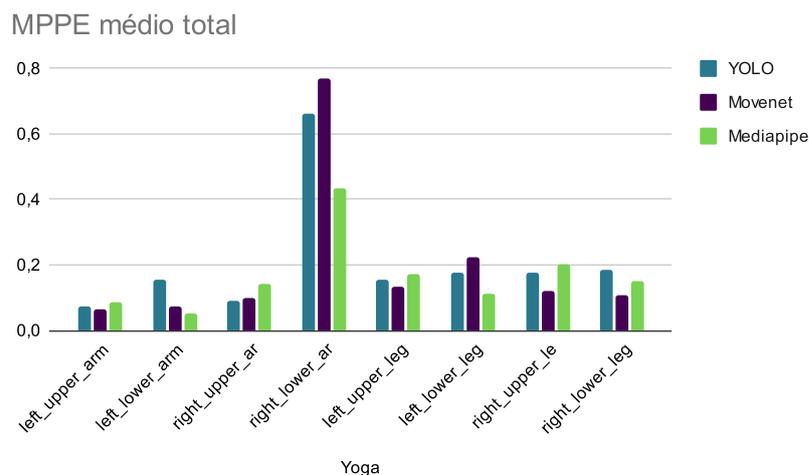
Fonte: Autor(2024).

Figura 36 – Médias para o $Erro_{OKS}$ no estudo de caso 1.



Fonte: Autor (2024).

Figura 37 – Médias para MPPE no estudo de caso 1.



Fonte: Autor (2024).

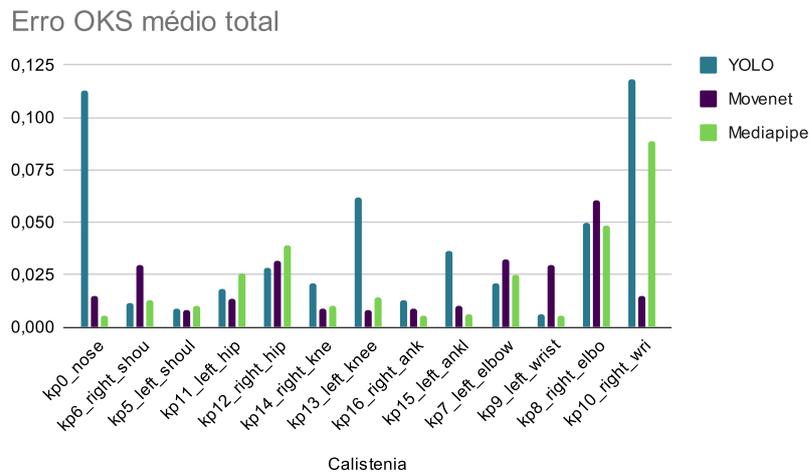
4.3 SÍNTESE DE DADOS

Levando em consideração os dados apresentados, é difícil selecionar imediatamente um modelo que se apresente como o mais adequado dado o balanço das métricas exploradas. Para isso foram gerados os gráficos das Figuras 36, 37, 38 e 39, que realizam uma média do desempenho dos modelos para cada keypoint ou parte.

Em primeiro lugar, na Figura 36 observa-se uma clara deficiência do Movenet em estimar corretamente o posicionamento de juntas inferiores, com destaque para o calcanhar direito, que foi um outlier apresentado na Seção 4.1.3.

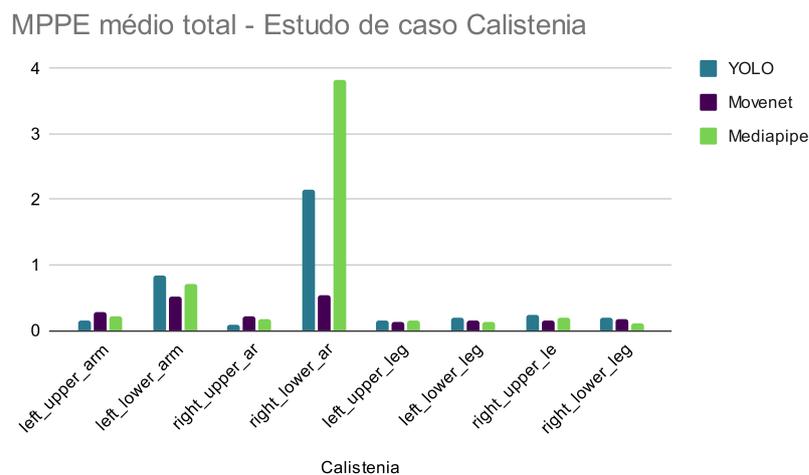
Na Figura 37 observa-se um gráfico de médias de MPPE para o estudo de

Figura 38 – Médias para o $Erro_{OKS}$ no estudo de caso 1.



Fonte: Autor (2024).

Figura 39 – Médias para MPPE no estudo de caso 1.



Fonte: Autor (2024).

caso 1, destaca-se um alto acúmulo de erro para o antebraço direito, liderado novamente pelo Movenet.

No entanto, ao tratar do estudo de caso 2, cujo $Erro_{OKS}$ está expresso na Figura 38, o cenário se inverte, com YOLO e Mediapipe apresentando altos valores, referentes principalmente aos dois casos críticos onde houve perda de keypoints por ocultação de juntas, detalhados nas Seções 4.2.1 e 4.2.2.

Observa-se na Figura 39, por fim, que MPPE médio para YOLO e Mediapipe foram muito acima da média no segundo estudo de caso novamente devido aos casos extremos explorados nas Seções 4.2.1 e 4.2.2.

5 CONCLUSÃO

Com base nos dados apresentados na Seção 4.3 e no Quadro 2, o MoveNet demonstrou ser o modelo mais equilibrado diante de cenários adversos. Apesar de apresentar o maior $Erro_{OKS}$ médio e um outlier significativo no estudo de caso 1, ele manteve os erros máximos de MPPE significativamente menores em comparação ao YOLO e ao Mediapipe, que apresentaram falhas mais graves, incluindo perda de pontos-chave.

No estudo de caso 2, o MoveNet teve o menor $Erro_{OKS}$ máximo, mesmo com baixa tolerância, evidenciando sua eficiência em posicionar keypoints em poses com juntas frequentemente ocultas. Assim, o MoveNet obteve o melhor desempenho geral, com menor variabilidade nos erros e maior consistência, destacando-se em cenários de calistenia. Contudo, os resultados indicam que melhorias são necessárias em todos os modelos para lidar com ângulos desafiadores, oclusões e poses complexas de forma mais robusta.

Os achados reforçam a importância da escolha cuidadosa de métricas para a avaliação de modelos de estimação de pose. Métricas como o OKS e o MPPE mostraram-se úteis, mas também expuseram limitações específicas em cenários de erro extremo, especialmente em casos de perda de keypoints ou inferências incorretas de elos.

O MoveNet também apresentou a menor utilização de processamento, conforme indicado no Quadro 1. Contudo, seu consumo de mais de 6 gigabytes de memória torna a tarefa altamente desafiadora em ambientes com capacidade computacional limitada. Situação semelhante foi observada com o YOLO, que exigiu quase 8 GB de memória. Esses desafios de desempenho destacam a necessidade de aprofundar as pesquisas voltadas para otimizações focadas em uso em dispositivos móveis.

Para trabalhos futuros, recomenda-se:

- Treinamento personalizado: Adaptar os modelos a contextos específicos, como exercícios físicos, utilizando conjuntos de dados mais variados e adequados às demandas de calistenia.
- Abordagens multicâmera: Explorar configurações que combinem múltiplas perspectivas para lidar com problemas de oclusão e melhorar a precisão da estimação de pose.
- Testes em plataformas móveis e embarcadas: Avaliar a viabilidade e o desempenho dos modelos em dispositivos de capacidade computacional limitada, garantindo aplicabilidade prática em cenários reais.
- Revisão sistemática de métricas: Ampliar a análise com a inclusão de métricas adicionais, como precisão e recall, particularmente em cenários de treinamento,

buscando uma avaliação mais abrangente e informativa.

Essas iniciativas podem contribuir para a evolução da estimação de pose humana, aumentando sua aplicabilidade em diferentes domínios e consolidando-a como uma ferramenta confiável em ambientes de uso real.

REFERÊNCIAS

- BADIOLA-BENGOA, A.; MENDEZ-ZORRILLA, A. A Systematic Review of the Application of Camera-Based Human Pose Estimation in the Field of Sport and Physical Exercise. *Sensors*, v. 21, n. 18, p. 5996, set. 2021. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/21/18/5996>.
- BAZAREVSKY, V. *et al.* **BlazePose: On-device Real-time Body Pose tracking**. arXiv, 2020. ArXiv:2006.10204 [cs]. Disponível em: <http://arxiv.org/abs/2006.10204>.
- CAO, Z.; HIDALGO, G.; SIMON, T.; WEI, S.-E.; SHEIKH, Y. **OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields**. arXiv, 2019. ArXiv:1812.08008 [cs]. Disponível em: <http://arxiv.org/abs/1812.08008>.
- CHANGOTRA, R. *et al.* A Comparative Study of Pose Estimation. *In: 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*. Gautam Buddha Nagar, India: IEEE, 2022. p. 210–216. ISBN 978-1-66546-643-1. Disponível em: <https://ieeexplore.ieee.org/document/9754071/>.
- CHEN, Y.; TIAN, Y.; HE, M. Monocular Human Pose Estimation: A Survey of Deep Learning-based Methods. *Computer Vision and Image Understanding*, v. 192, p. 102897, mar. 2020. ISSN 10773142. ArXiv:2006.01423 [cs]. Disponível em: <http://arxiv.org/abs/2006.01423>.
- DAWN, K. **Build an AI Fitness Trainer Using MediaPipe For Squat Analysis — learnopencv.com**. 2022. Disponível em: <https://learnopencv.com/ai-fitness-trainer-using-mediapipe/>.
- DUBEY, S.; DIXIT, M. A comprehensive survey on human pose estimation approaches. *Multimedia Systems*, v. 29, n. 1, p. 167–195, fev. 2023. ISSN 0942-4962, 1432-1882. Disponível em: <https://link.springer.com/10.1007/s00530-022-00980-0>.
- DUTTA, T.; DAWN, K. **Object Keypoint Similarity in Keypoint Detection**. 2023. Acessado em: 2024-10-30. Disponível em: <https://learnopencv.com/object-keypoint-similarity/>.
- JOSYULA, R.; OSTADABBAS, S. **A Review on Human Pose Estimation**. arXiv, 2021. ArXiv:2110.06877 [cs]. Disponível em: <http://arxiv.org/abs/2110.06877>.
- MAJI, D.; NAGORI, S.; MATHEW, M.; PODDAR, D. *In: YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss*. New Orleans, LA, USA: IEEE, 2022. p. 2636–2645. ISBN 978-1-66548-739-9. Disponível em: <https://ieeexplore.ieee.org/document/9857215/>.
- NINA, A. **GitHub - arthurhn/HPE_models — github.com**. 2024. Disponível em: https://github.com/arthurhn/HPE_models.
- NINA, A. **GitHub - arthurhn/Performance_tests: Final clean version of performance tests on HPE models — github.com**. 2024. Disponível em: https://github.com/arthurhn/Performance_tests.

OVED, D.; ALVARADO, I.; GALLO, A. **Real-time Human Pose Estimation in the Browser with TensorFlow.js** — blog.tensorflow.org. 2018. Disponível em: <https://blog.tensorflow.org/2018/05/real-time-human-pose-estimation-in.html>.

RICHTER, F. **Charted: There Are More Phones Than People in the World**. 2023. Acessado em: 2024-10-30. Disponível em: <https://www.weforum.org/stories/2023/04/charted-there-are-more-phones-than-people-in-the-world/>.

ROHAN, A.; RABAH, M.; HOSNY, T.; KIM, S.-H. Human Pose Estimation-Based Real-Time Gait Analysis Using Convolutional Neural Network. **IEEE Access**, v. 8, p. 191542–191550, 2020. ISSN 2169-3536. Disponível em: <https://ieeexplore.ieee.org/document/9220146/>.

SOUZA, G. L.; MOREIRA, N. B.; CAMPOS, W. Ocorrência e Características de Lesões entre Praticantes de Musculação. *Saúde e Pesquisa*, v. 8, n. 3, p. 469, dez. 2015. ISSN 2176-9206, 1983-1870. Disponível em: <http://periodicos.unicesumar.edu.br/index.php/saudpesq/article/view/4476>.

SUGAWARA, E. K. Estimativa de Postura: Uma abordagem multicâmera para classificação e avaliação de exercícios físicos. UFSC, 2022.

TENSORFLOW. **Pose Detection Tensorflow.js models**. 2022. Disponível em: <https://github.com/tensorflow/tfjs-models/blob/master/pose-detection/README.md>.

VOTEL, R.; LI, N. **Next-Generation Pose Detection with MoveNet and TensorFlow.js** — blog.tensorflow.org. 2021. Disponível em: <https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html>.

**APÊNDICE A – TABELA DE MÉDIAS DE OKS POR KEYPOINT POR MODELO
PARA AS POSES DE YOGA**

Tabela 2 – Médias de OKS_{var} por ponto-chave.

Ponto-chave	YOLO	Mediapipe	Movenet
$kp0_{nose}$	0,999933	0,999430	0,999759
$kp6_{rightshoulder}$	0,999661	0,999837	0,999880
$kp5_{leftshoulder}$	0,999814	0,999866	0,999871
$kp11_{leftthip}$	0,999416	0,999362	0,999467
$kp12_{rightthip}$	0,999561	0,999258	0,999654
$kp14_{rightknee}$	0,999108	0,998041	0,996926
$kp13_{leftknee}$	0,997783	0,998417	0,998515
$kp16_{rightankle}$	0,998260	0,995935	0,984486
$kp15_{leftankle}$	0,997293	0,999300	0,999291
$kp7_{leftelbow}$	0,999860	0,999857	0,999900
$kp9_{leftwrist}$	0,999841	0,999487	0,999773
$kp8_{rightelbow}$	0,999720	0,999864	0,999944
$kp10_{rightwrist}$	0,999798	0,999906	0,999616

Fonte: Autor (2024).

**APÊNDICE B – TABELA DE MÉDIAS DE OKS POR KEYPOINT POR MODELO
PARA AS POSES DE YOGA**

Tabela 3 – Médias de $OKS_{0.1}$ por ponto-chave.

Ponto-chave	YOLO	Mediapipe	Movenet
$kp0_{nose}$	0,999545	0,996189	0,998374
$kp6_{rightshoulder}$	0,979426	0,989934	0,992603
$kp5_{leftshoulder}$	0,988564	0,991795	0,992003
$kp11_{leftthip}$	0,965826	0,961454	0,967746
$kp12_{rightthip}$	0,973299	0,956516	0,978868
$kp14_{rightknee}$	0,957060	0,943462	0,881055
$kp13_{leftknee}$	0,903317	0,940276	0,946981
$kp16_{rightankle}$	0,944152	0,934726	0,739986
$kp15_{leftankle}$	0,914621	0,974560	0,974717
$kp7_{leftelbow}$	0,992812	0,992711	0,994891
$kp9_{leftwrist}$	0,993919	0,981821	0,991365
$kp8_{rightelbow}$	0,985672	0,993034	0,997154
$kp10_{rightwrist}$	0,992369	0,996422	0,985650

Fonte: Autor (2024).