



Ferramentas de Classificação e Reconhecimento de Imagem para o Aprendizado de Artes Marciais

Aleksander Yacovenco

JUIZ DE FORA
MARÇO, 2021

Ferramentas de Classificação e Reconhecimento de Imagem para o Aprendizado de Artes Marciais

ALEKSANDER YACOVENCO

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientador: Marcelo Caniato Renhe

JUIZ DE FORA
MARÇO, 2021

FERRAMENTAS DE CLASSIFICAÇÃO E RECONHECIMENTO DE IMAGEM PARA O APRENDIZADO DE ARTES MARCIAIS

Aleksander Yacovenco

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Marcelo Caniato Renhe
Doutor em Engenharia de Sistemas e Computação

Rodrigo Luis de Souza da Silva
Doutor em Engenharia Civil

Luiz Maurílio da Silva Maciel
Doutor em Engenharia de Sistemas e Computação

JUIZ DE FORA
12 DE MARÇO, 2021

Resumo

Ao longo dos últimos anos, pôde ser percebida a popularização de ferramentas de reconhecimento e classificação de imagens e de movimento (mais notoriamente Redes Neurais Artificiais) em diversas pesquisas relacionadas ao movimento humano. Dispositivos de captura de imagem ou de movimento (como câmeras, acelerômetros e giroscópios), juntamente com ferramentas de reconhecimento e classificação, são usados em vários tipos de estudo, como analisar e otimizar o treino de atletas de alta performance, avaliar a reabilitação de pessoas após acidentes ou categorizar diferentes tipos de movimento cotidiano. Usando tais ferramentas, essa pesquisa tem como objetivo classificar e avaliar determinadas posturas de artes marciais, mais especificamente HEMA (*Historical European Martial Arts*, ou Artes Marciais Históricas Europeias, em português), servindo essencialmente como uma ferramenta para o treino autônomo. A partir dos dados coletados, pretende-se averiguar se tal ferramenta pode de fato servir o propósito de auxiliar praticantes autodidatas, ao menos nos aspectos mais básicos da arte marcial em questão.

Palavras-chave: Classificação de imagem, detecção de imagem, artes marciais, *Open-Pose*, HEMA.

Abstract

For the past few years, there's been a growth in popularity of tools for recognizing and classifying images and movement (most notably Artificial Neural Networks) in many researches related to human motion. Image and movement capture devices (like cameras, accelerometers and gyroscopes), along with tools for recognition and classification, are used in various kinds of researches, like analyzing and optimizing the train routine of high performance athletes, evaluating the rehabilitation of people after accidents or categorizing different types of everyday natural human movements. By using such tools, this research has the goal to classify and evaluate a few set postures of martial arts, more specifically HEMA (Historical European Martial Arts), essentially serving as an autonomous training tool. From the collected data, it's possible to verify the efficacy of this tool and if it can in fact help learning some of the most basic aspects of the set martial art autonomously.

Keywords: Image classification, image detection, martial arts, HEMA, OpenPose.

Agradecimentos

Em primeiro lugar devo agradecer ao meu orientador, Marcelo Caniato, por sua paciência e suas valiosas observações. Graças a ele fui capaz de concluir este trabalho e crescer como pesquisador. De maneira similar, agradeço a professora Luciana Campos, minha primeira orientadora de Iniciação Científica, por me auxiliar na minha primeira jornada pelo meio acadêmico.

Agradeço também a Iluska Coutinho e Jorge Felz, que sempre estavam lá por mim quando eu precisava. Tenho certeza de que me adaptar a Juiz de Fora, uma cidade completamente nova pra mim na época, teria sido infinitamente mais difícil sem a presença deles, e sem eles eu não teria tantas memórias agradáveis quanto tenho hoje.

Para desfrutar dos momentos agradáveis assim como suportar os difíceis foi essencial a presença de minha namorada Laura Helena Amorim, assim como a presença de meus amigos, dentre os quais se destacam Rafael Terra, Mateus Marim, João Gabriel Malaguti, Luiz Alexandre Alves e Caio Pedroso.

Por fim gostaria de agradecer ao meu avô Jobel Teixeira, que sempre acreditou em mim. Sem o seu inabalável suporte, creio que sequer teria entrado no curso de Ciência da Computação. Sua presença sempre foi motivadora na minha vida, e por isso sou eternamente grato.

*“Not the easiest climb, is it? But I guess
that’s what I was looking for...”*

Madeline (Celeste, 2018)

Conteúdo

Lista de Figuras	6
Lista de Tabelas	7
Lista de Abreviações	8
1 Introdução	9
1.1 Trabalhos relacionados	11
2 Embasamento teórico	15
2.1 Redes Neurais Artificiais e Visão Computacional	15
2.2 <i>OpenPose</i> e <i>TensorFlow</i>	18
2.3 O corpo humano e o modelo tipo esqueleto	19
2.4 Reconhecimento e classificação de imagens	20
2.5 A arte marcial	22
3 Desenvolvimento	26
3.1 As posturas	29
3.2 Orientação dos pés e calcanhares	29
3.3 Detecção da arma	31
3.4 Correção da inferência	32
4 Análise de dados	33
4.1 Limitação da base de dados	38
5 Conclusão	39
Referências Bibliográficas	41

Lista de Figuras

1.1	Exemplos de dispositivos analógicos.	13
2.1	Modelo do neurônio biológico (Haykin, 2007).	16
2.2	Modelo do neurônio artificial (Haykin, 2007).	16
2.3	<i>Perceptron</i> de múltiplas camadas com duas camadas ocultas e um neurônio de saída (Haykin, 2007).	17
2.4	Módulo de reconstrução tridimensional do <i>OpenPose</i> (Cao et al, 2019).	19
2.5	Modelos de esqueleto (Cao et al, 2019).	20
2.6	Representação das mãos no <i>OpenPose</i> (Smith et al, 2017).	22
2.7	Ilustrações restauradas e coloridas das 4 posturas básicas (Wiktenauer, 2020).	23
3.1	Desambiguação da postura de pernas.	28
3.2	Os 5 ângulos de captura. Imagens extraídas de Rüter (2016).	29
3.3	Interseção das linhas dos pés para avaliação da corretude da base (módulo de detecção de objetos desligado).	31
4.1	Funcionamento da ferramenta.	34
4.2	Taxa de acerto da ferramenta sem a detecção de objetos e sem regressão linear.	35
4.3	Taxa de acerto com e sem detecção de objetos (e sem regressão linear).	36
4.4	Taxa de acerto com e sem regressão linear (sem detecção de objetos).	37

Lista de Tabelas

2.1	As 4 guardas principais segundo Meyer (1570). Figuras disponíveis em Wiktenauer (2020).	24
2.2	Posições dos pés em relação à linha central (BoxingIsOn, 2021).	25
3.1	Primeira abstração das posturas.	30

Lista de Abreviações

API *Application Programming Interface*

HEMA *Historical European Martial Arts*

RNA Rede Neural Artificial

1 Introdução

Ferramentas de classificação de movimento vêm sendo usadas extensivamente ao longo dos últimos anos em diversas pesquisas, e o seu uso apresenta resultados significativos para analisar e otimizar o treino de determinadas atividades físicas (Worsey et al, 2019; He et al, 2018). Os métodos para capturar o movimento humano variam entre cada trabalho. Alguns utilizam processamento de imagens e Redes Neurais Artificiais (RNAs) treinadas para trabalhar com Visão Computacional (Farias, 2012), outros utilizam tecnologias como o Kinect (Kamaya et al, 2018), e há aqueles que utilizam outros sensores de movimento posicionados estrategicamente pelo corpo da pessoa como acessórios (Zhang et al, 2017; Chen et al, 2017; Barrantes et al, 2017). Os objetivos das pesquisas também variam consideravelmente, mas todos remetem à classificação de movimento humano (Reyes-Ortiz et al, 2016; Shoab et al, 2016, 2014; Wu et al, 2012).

Dentre as atividades físicas contempladas por pesquisas nessa área, as artes marciais constituem um grande subconjunto, com revisões bibliográficas detalhadas e bem elaboradas (Zhang et al, 2017; Kamaya et al, 2018). Resumidamente, os objetivos dos trabalhos descritos por essas revisões geralmente consistem em aprimorar algum aspecto do treinamento das artes marciais praticadas. Mas para entender o que aprimorar, primeiro é necessário entender o contexto.

Como qualquer atividade física, toda arte marcial precisa de um ambiente adequado para ser praticada. Em primeiro lugar, é necessário um ou mais instrutores ou, minimamente, um material didático através do qual possa ser assimilado o conteúdo da arte. Em segundo lugar, é necessário um espaço suficientemente amplo para praticar os movimentos. No caso de artes marciais que utilizam armas brancas (espadas e bastões, por exemplo), esse espaço amplo torna-se ainda mais necessário. Por fim, é necessário um ou mais parceiros de treino com quem se pode praticar os movimentos, validar a sua eficácia e possivelmente se adaptar a diferentes tipos físicos (é de conhecimento comum que são necessários pequenos ajustes na técnica ao lutar contra pessoas de diferentes tamanhos e pesos).

A dificuldade de atender esses requisitos é ainda mais acentuada quando se quer treinar alguma arte não popularizada, como HEMA. A sigla HEMA significa *Historical European Martial Arts* e refere-se ao conjunto de artes marciais praticadas na Europa entre os séculos XIII e XVIII, aproximadamente. Hoje, praticantes de HEMA tentam reconstruir essas artes marciais com base em registros históricos, em sua maioria disponíveis em museus e grandes bibliotecas nacionais. Atualmente, há poucos centros que treinam HEMA, e esses poucos centros existentes estão concentrados nos EUA, no Canadá e em diversos países da Europa. Além da falta de disponibilidade de centros ou academias com instrutores certificados, há a dificuldade financeira. Artes marciais com armas brancas são caras, e os equipamentos de HEMA são semelhantes aos da esgrima olímpica: espada, máscara de esgrima, luvas com proteção especial, jaquetas de esgrima, caneleiras, joelheiras e outros itens opcionais. Ainda que o praticante tenha o equipamento mínimo necessário para começar a praticar, normalmente é difícil encontrar um parceiro de treino e um local para treinar rotineiramente.

A hipótese deste trabalho é a de que os aspectos mais básicos da arte marcial podem ser assimilados através de uma ferramenta virtual, sem necessidade de um instrutor. Um ambiente virtual solucionaria grande parte das dificuldades iniciais, ainda que temporariamente. O custo financeiro seria minimizado, eliminando a necessidade de pagar pelo equipamento de proteção ou pelo acesso a uma academia. Além disso, não haveria a necessidade de um instrutor, dado que a ferramenta possuiria uma base de dados extensa o bastante para validar a postura do praticante, permitindo a este praticar os movimentos sem sair de casa.

O objetivo principal é avaliar a aplicabilidade de ferramentas de reconhecimento e classificação de imagens no reconhecimento de determinadas posturas de HEMA, utilizando dispositivos simples de captura de imagem (câmeras), de modo a permitir a um praticante verificar a corretude de sua postura. Assim, pretende-se desenvolver um sistema capaz de avaliar posturas humanas em imagens 2D através do uso das bibliotecas *OpenPose* (Cao et al, 2019) e *TensorFlow* (Abadi et al, 2015). Essas bibliotecas adquirem os pontos de articulação das pessoas na imagem com base em modelos pré treinados (COCO e BODY_25). Além disso, objetiva-se obter experimentalmente níveis aceitáveis

de variação para cada postura, dada uma base de dados processada pela ferramenta, ou seja, a diferença numérica máxima entre uma dada postura e a postura análoga presente na base de dados, de forma que a postura analisada ainda possa ser considerada correta.

Para validar as técnicas e posturas capturadas, o material base utilizado é o livro *The Art of Combat* (Forgeng, 2015). O livro é uma tradução do manual *Gründliche Beschreibung der Kunst des Fechtens* (Meyer, 1570), cuja cópia certificada usada para transcrição e tradução encontra-se hoje no *Metropolitan Museum of Art*, em Nova York, nos Estados Unidos. A ferramenta também é alimentada com uma base de dados de imagens e quadros de vídeos de instrutores certificados, para garantir que todos os movimentos estejam biomecanicamente corretos. Com o objetivo principal cumprido, o próximo passo seria avaliar as posturas do praticante e exibir instruções de como ele as poderia corrigir.

1.1 Trabalhos relacionados

Tecnologias de detecção e captura de movimento não são tão recentes quanto se imagina. Dispositivos analógicos e mecânicos para capturar o movimento humano datam de antes do surgimento da computação como é conhecida hoje, como por exemplo as patentes de medidores do vetor de força de tacadas de golfe (Nicholas, 1965; Evans, 1966; White, 1929; Edward, 1965). Hoje há duas formas mais populares para capturar movimentos: dispositivos analógicos e imagens adquiridas com câmeras.

Uma pesquisa relevante é a de Robert-Lachaine et al (2020). Seu objetivo é validar novos dispositivos analógicos disponíveis no mercado, financeiramente mais acessíveis mas não garantidamente precisos o bastante. Estes dispositivos são capazes de adquirir os valores de aceleração e velocidade de rotação de movimentos. O experimento é realizado da seguinte forma: cinco adultos saudáveis realizam a mesma atividade (levantar, mover e abaixar caixas) em uma mesma janela de tempo de aproximadamente meia hora. Todos os participantes vestem conjuntos de sensores de movimento e no fim do experimento os resultados são comparados com o que foi obtido através da captura óptica. Isso pode gerar uma dúvida: se a captura óptica é tão mais precisa a ponto de ser usada como referência, por que então usar dispositivos analógicos? Algumas razões são a facilidade

de capturar movimentos em 3D sem problemas como oclusão ou a falta de uma linha guia, menor custo de processamento e, conseqüentemente, maior facilidade de apresentar os dados em tempo real.

Entretanto, dispositivos analógicos apresentam certas desvantagens. A mais crucial é que, apesar de apresentarem bons resultados ao medir movimentos, eles não são tão eficientes para medir posturas estáticas. Isso pode ser visto nas revisões de Moeslund et al (2001) e He et al (2018), ou comparando as revisões bibliográficas de Chen et al (2020) e Worsey et al (2019), respectivamente sobre tecnologias de Visão Computacional e de sensores analógicos. Isso é uma conclusão lógica, visto que acelerômetros e giroscópios são feitos para adquirir os vetores de aceleração e rotação do dispositivo, respectivamente. Assim, quando a pessoa vestindo o dispositivo está parada, o acelerômetro computa apenas a força da gravidade (o que muda é o sentido do vetor, que depende da orientação do dispositivo no espaço, visto que a força da gravidade aponta sempre para baixo independente dessa orientação) e o giroscópio registra uma rotação nula. Existem também outros dispositivos analógicos além de acelerômetros e giroscópios, como dispositivos mecânicos, mas eles apresentam desvantagens muito similares. Em contraste, dados capturados com *softwares* de Visão Computacional que trabalham com uma imagem de cada vez independem do movimento e, portanto, são mais apropriados para trabalhar com posturas estáticas.

Outra pesquisa que examina a diferença entre esses métodos é a de Polak et al (2016). Polak estuda as vantagens e desvantagens de métodos computacionais de análise de técnicas de artes marciais e aponta, dentre outras, duas desvantagens consideráveis de sensores analógicos: o alto custo na época de sua pesquisa, que permanece alto atualmente, e a acessibilidade. Apenas pesquisadores usam dispositivos analógicos para medir os movimentos dos praticantes. Tais dispositivos não fazem parte do dia a dia dos instrutores e praticantes, e apenas vestir tais sensores não é uma tarefa fácil de tornar cotidiana, como pode ser visto na Figura 1.1. Câmeras, por outro lado, são consideravelmente menos trabalhosas, sendo mais fácil incluir um *software* na rotina de treino como ferramentas para o instrutor, melhorando assim sua análise dos movimentos dos praticantes. Isso foi feito, por exemplo, no trabalho de Cynarski et al (2014), onde é criado um modelo

matemático de praticantes de Karatê com base em dados adquiridos a partir de câmeras.



(a) Dispositivos usados na pesquisa de Vences Brito et al (2011).



(b) Imagem usada como exemplo na pesquisa de Polak et al (2016).

Figura 1.1: Exemplos de dispositivos analógicos.

Como pode ser visto na revisão de Chen et al (2020), há diversos métodos para aquisição dos dados sobre o corpo humano a partir de imagens monoculares. O *OpenPose* (Cao et al, 2019) e o *TensorFlow* (Abadi et al, 2015), entretanto, já possuem implementações de alguns desses métodos para adquirir os dados no formato de esqueleto (onde os pontos-chave são as juntas, como joelhos, ombros e cotovelos), mais especificamente nos modelos COCO e BODY_25.

Vistas as tecnologias de captura de imagem e movimento, ainda há a preocupação em como analisar academicamente as nuances da arte marcial em questão. O livro de Forggeng (2015), tradução do manuscrito de Meyer (1570), cobre os pontos necessários para o aprendizado da arte marcial, mas não para a parametrização dos movimentos. O trabalho de Merkert (2019), entretanto, trata justamente sobre isso. O autor define o que é HEMA e discorre sobre as interpretações modernas das técnicas, tanto em ambientes de prática ou treino quanto em ambientes de *freefight*, que são duelos entre duas pessoas com armas não afiadas e usando equipamento completo de proteção similar ao de esgrima olímpica. Em sua tese, Merkert encontra as diferenças de execução das técnicas nesses diferentes ambientes. Entretanto, é um praticante experiente, e não uma máquina, que classifica as técnicas para posterior análise dos resultados. Apesar de Merkert trabalhar com acelerômetros e medir principalmente a velocidade dos movimentos, o que não é muito útil em contextos em que o objetivo é medir posturas estáticas, seus dados são robustos e significativos para qualquer pesquisa relacionada a HEMA.

2 Embasamento teórico

Por se tratar de um estudo multidisciplinar, o embasamento teórico não abrange apenas a área da Ciência da Computação, mas também a área da Saúde, em especial alguns conceitos de Biomecânica. O trabalho de Merkert (2019) aborda especificamente a biomecânica dos movimentos da escola de esgrima alemã de Liechtenauer, que é a mesma tradição seguida por Meyer (1570). Alguns desses conceitos serão apresentados neste capítulo, além da parte teórica associada ao reconhecimento e classificação de imagens.

2.1 Redes Neurais Artificiais e Visão Computacional

Redes Neurais Artificiais, ou RNAs, são modelos computacionais especializados em reconhecimento de padrões e inspirados em redes neurais biológicas (Kovács, 2002). Este modelo nada mais é que uma representação matemática das conexões entre os neurônios biológicos. Simplificadamente, os dendritos recebem impulsos ou potenciais de ação de sinapses, que são transmitidos ao corpo do neurônio, que então envia os impulsos recebidos para o axônio, como mostrado na Figura 2.1:

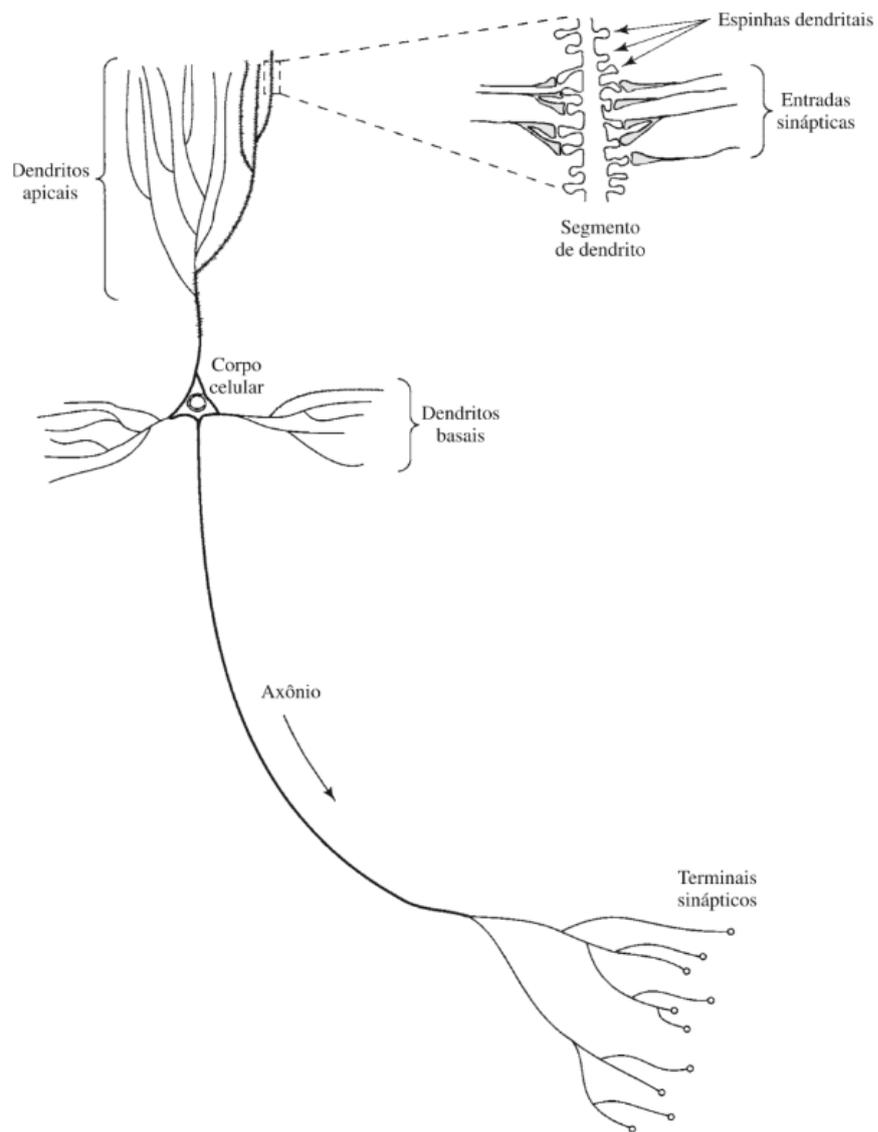


Figura 2.1: Modelo do neurônio biológico (Haykin, 2007).

Essas características podem ser expressas matematicamente da forma indicada pela Figura 2.2.

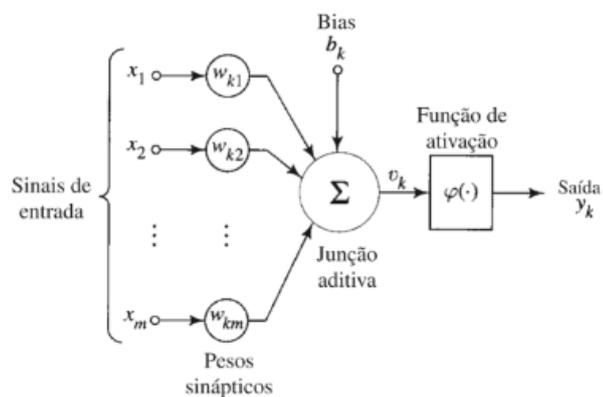


Figura 2.2: Modelo do neurônio artificial (Haykin, 2007).

A RNA recebe então várias entradas que são processadas por cada neurônio, e estes neurônios produzem uma saída que pode ser a saída final ou pode ser passada como entrada para outra camada de neurônios, como pode ser visto na Figura 2.3.

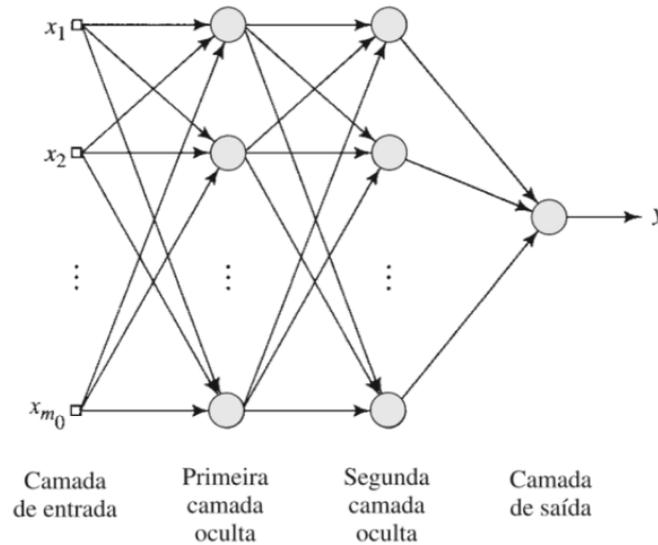


Figura 2.3: *Perceptron* de múltiplas camadas com duas camadas ocultas e um neurônio de saída (Haykin, 2007).

Os pesos sinápticos são ajustados pela própria RNA durante um processo de treinamento definido por um algoritmo, etapa na qual a RNA recebe entradas rotuladas. Após o treino, caso tudo tenha ocorrido de acordo com o esperado, a RNA está apta a receber entradas inéditas e as classificar com base nos dados já classificados no treinamento.

RNAs são usadas amplamente na área de Visão Computacional. A Visão Computacional nada mais é, afinal, que uma máquina extraíndo informações importantes de uma imagem ou de um conjunto de imagens (Milano et al, 2014). Pode-se concluir, então, que RNAs são ideais para classificar novos dados a partir de dados similares processados anteriormente, o que encaixa perfeitamente nesse contexto. Entretanto, as RNAs não são as únicas tecnologias atuantes na Visão Computacional. Também pode ser realizado um pré-processamento e a extração de determinadas características da imagem, como textura e bordas. Em alguns casos, a própria RNA ajusta os pesos sinápticos para realizar essa etapa de pré processamento da imagem, e somente depois passa para a etapa de classificação.

2.2 *OpenPose* e *TensorFlow*

O *OpenPose* (Cao et al, 2019) e o *TensorFlow* (Abadi et al, 2015) são plataformas que usam RNAs e algoritmos de classificação e são capazes de extrair de uma imagem um modelo matemático do corpo humano (caso um humano esteja presente na imagem). Embora, nesse contexto, ambas sejam usadas essencialmente para a mesma tarefa, há algumas diferenças significativas entre elas.

O *TensorFlow* é uma plataforma mais completa e generalista para aplicações de *Machine Learning*, ou Aprendizado de Máquina. Seu foco principal não é a classificação de posturas humanas, apesar de ele possuir uma API (*Application Programming Interface*, ou Interface para Programação de Aplicação) para tal. Dentre sua grande variedade de APIs e aplicações prontas, destacam-se o módulo de detecção do corpo humano e o módulo de detecção de objetos, que permitem treinar uma RNA especializada para a classificação de imagens com uma base de dados específica.

Já o *OpenPose* é uma plataforma especializada em detecção do corpo humano, com módulos de detecção tanto bidimensionais quanto tridimensionais e diversos códigos de exemplo. O módulo de detecção tridimensional, entretanto, requer mais de uma câmera conectada à máquina, cada uma filmando a mesma pessoa ao mesmo tempo, bem como um plano com quadrados pretos e brancos desenhados na forma de um tabuleiro de xadrez, de forma a facilitar a detecção do ângulo e a calibração da câmera. Tais recursos não estavam disponíveis para esta pesquisa, portanto este módulo não foi utilizado. Essa funcionalidade pode ser visualizada mais facilmente com a Figura 2.4.

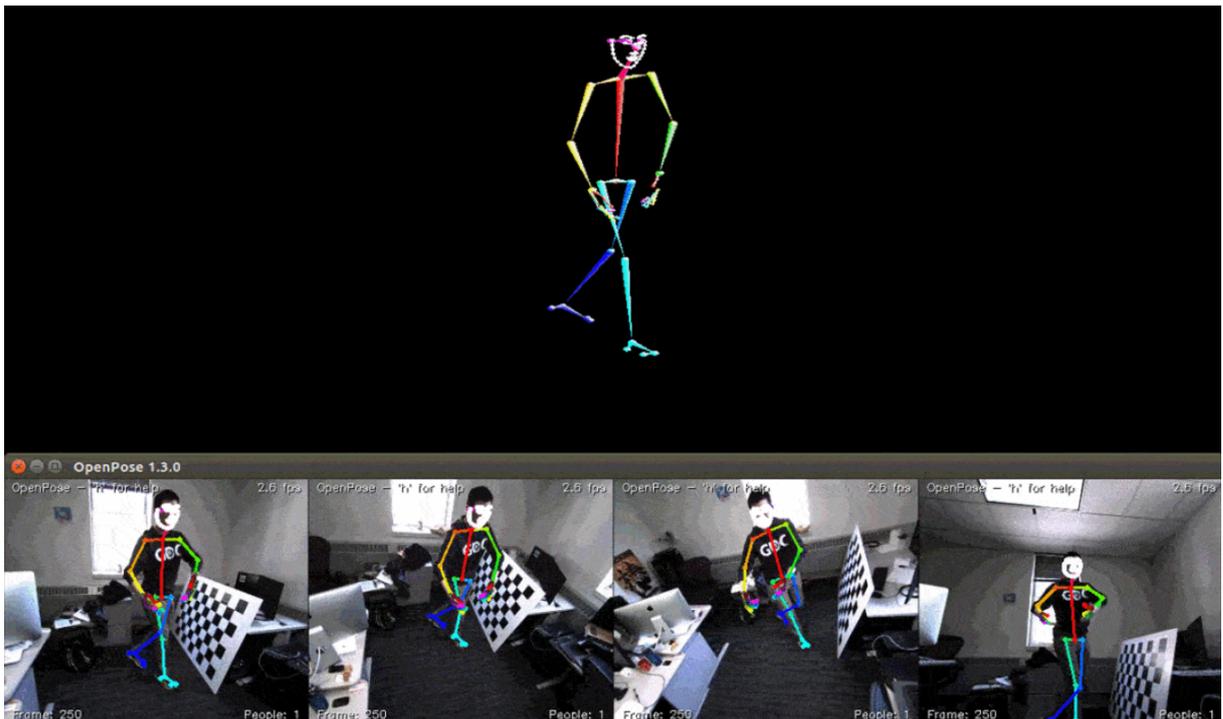


Figura 2.4: Módulo de reconstrução tridimensional do *OpenPose* (Cao et al, 2019).

2.3 O corpo humano e o modelo tipo esqueleto

O *OpenPose* e o *TensorFlow* trabalham com o modelo chamado de “esqueleto”. Existem 3 modelos do corpo humano considerados os mais comuns na literatura, que são o de esqueleto, o de contorno e o de volume (Chen et al, 2020). Os modelos de contorno e volume são mais apropriados para representações tridimensionais, entretanto.

O modelo de esqueleto é essencialmente um grafo, representando o corpo humano como um conjunto de vértices que são as juntas e um conjunto de arestas que são os membros. A maior desvantagem do modelo de esqueleto é que ele não representa todas as informações do corpo, como textura, volume dos membros etc. Porém, essas informações são irrelevantes para este estudo, visto que é necessário apenas saber a posição e rotação dos membros no espaço.

Tendo isso em vista, o modelo esqueleto é ideal para a tarefa em questão. As bibliotecas *OpenPose* e *TensorFlow* possuem implementações para adquirir o modelo a partir de uma imagem ou vídeo. O *TensorFlow* usa o modelo COCO, enquanto o *OpenPose* pode usar tanto o COCO quanto o BODY_25, que é um modelo proposto para ser mais completo. Como pode ser notado nas imagens da Figura 2.5, ele possui mais

pontos de articulação para os pés, enquanto o modelo COCO termina nos calcanhares. Esses pontos são extremamente importantes para avaliar as posturas posteriormente.

A princípio, a ausência no modelo COCO de alguns dos pontos de articulação presentes no modelo BODY_25 seriam o bastante para descartar o *TensorFlow*. No entanto, o módulo de detecção de objetos do *TensorFlow*, quando usado em conjunto com o *OpenPose*, pode aumentar a precisão das inferências, conforme será visto mais adiante.

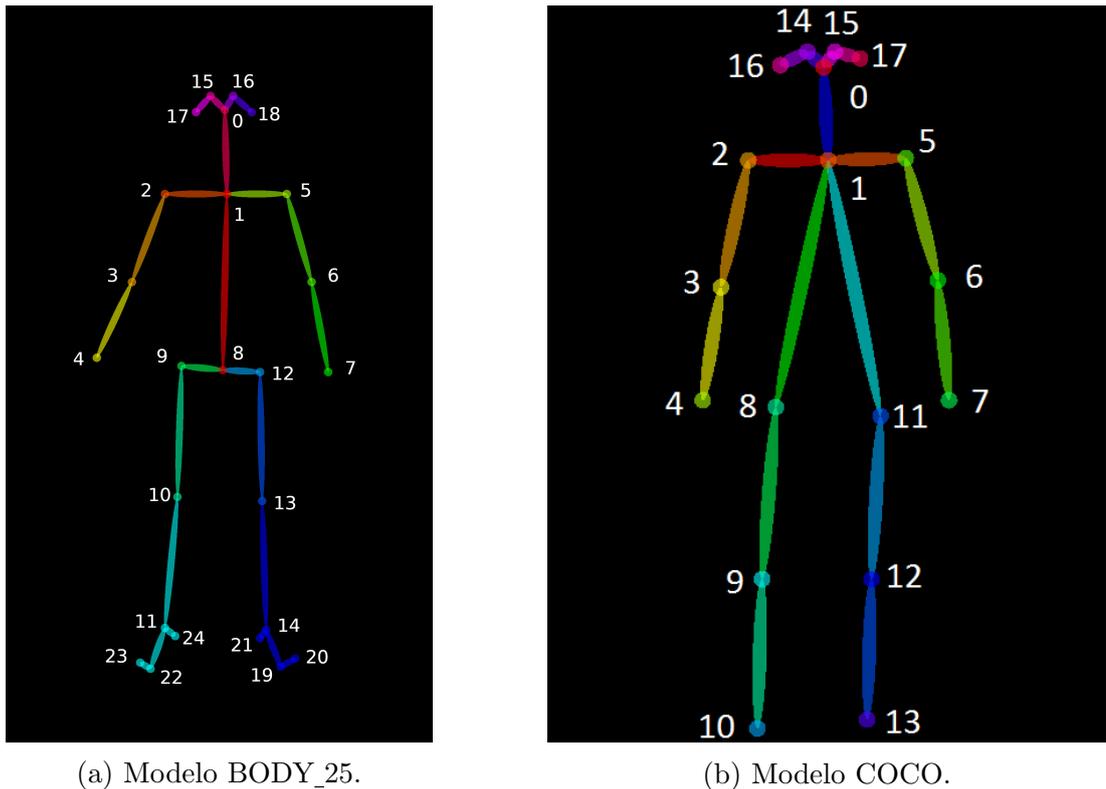


Figura 2.5: Modelos de esqueleto (Cao et al, 2019).

2.4 Reconhecimento e classificação de imagens

Existem 3 módulos importantes para este estudo:

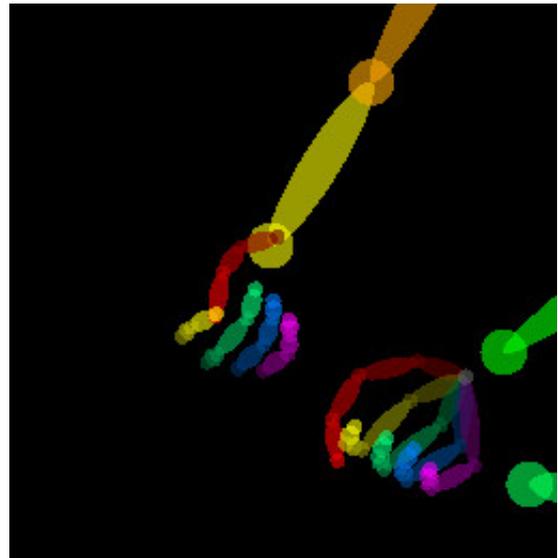
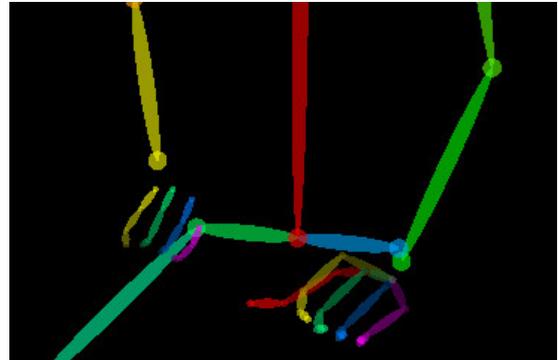
- Detecção de corpo: adquire o modelo de esqueleto para uma dada imagem contendo uma ou mais pessoas;
- Detecção de mão: similar à detecção de corpo, mas a RNA é treinada especificamente para reconhecer mãos em vez de um corpo inteiro. Este módulo é mais difícil de implementar pois as mãos geralmente possuem uma resolução muito menor que o resto do corpo. Logo, grande parte das pesquisas utiliza essa ferramenta apenas

quando apenas as mãos são relevantes para alguma determinada finalidade como, por exemplo, na pesquisa de Bansal (Bansal, 2020), em que a RNA é usada para traduzir a linguagem de sinais estadunidense;

- Detecção de objetos: adquire a posição de um objeto em uma imagem. Apesar de já haver bases de dados disponíveis *online* para RNAs de reconhecimento de objetos, geralmente a RNA deve ser treinada para reconhecer algum novo objeto específico.

Os módulos de detecção de mãos e de objetos são menos importantes que o módulo de detecção de corpo, mas também são extremamente úteis, visto que não é possível medir a rotação dos pulsos e a posição das mãos usando apenas o módulo principal. Para artes marciais que não fazem uso de armas brancas, isso pode ser uma funcionalidade opcional, mas, em uma arte marcial onde há armas envolvidas, a posição da arma no espaço e a posição das mãos que seguram a arma tornam-se extremamente relevantes.

Para exemplificar, a Figura 2.6 apresenta exemplos de mãos capturadas e processadas pelo *OpenPose*, nas empunhaduras *hammer grip* ou “pegada martelo”, em que o praticante segura a espada com o punho fechado, assim como um martelo, e *thumb grip* ou “pegada do dedão”, em que o praticante posiciona o dedo sobre a parte chata da lâmina para ganhar maior estabilidade nos movimentos de rotação nesse plano:

(a) *Hammer grip*, imagem original.(b) *Hammer grip*, imagem processada.(c) *Thumb grip*, imagem original.(d) *Thumb grip*, imagem processada.Figura 2.6: Representação das mãos no *OpenPose* (Smith et al, 2017).

2.5 A arte marcial

De acordo com Meyer (Meyer, 1570), a esgrima alemã da tradição Liechtenauer possui 4 posturas, ou guardas, básicas:

- *Alber*, ou “Tolo”;
- *Ochs*, ou “Touro”;
- *Pflug*, ou “Arado”; e
- *Vom Tag*, ou “do Dia”, também chamada de “*High Guard*” ou “Guarda Alta”.

Vale lembrar que as guardas *Ochs* e *Pflug* possuem variações para a esquerda ou para a direita, trocando a perna da frente e espelhando a espada horizontalmente, mas

mantendo as mãos no mesmo ponto da empunhadura da espada (um erro comum é trocar as mãos para não precisar cruzar os braços ao trocar de *Ochs* à esquerda para *Ochs* à direita).



(a) *Vom Tag* (esquerda) e *Alber* (direita).



(b) *Ochs* à direita (esquerda) e *Pflug* à direita (direita).

Figura 2.7: Ilustrações restauradas e coloridas das 4 posturas básicas (Wiktenauer, 2020).

Convenientemente, Meyer detalhou as guardas da esgrima alemã minuciosamente

em seu livro, *Gründtliche Beschreibung der Kunst des Fechtens* (Meyer, 1570), algo inédito para sua época. Uma breve leitura do capítulo sobre as guardas da tradução de sua obra, *The Art of Combat: A German Martial Arts Treatise of 1570* (Forgeng, 2015) é o suficiente para esclarecer essa ideia. De forma resumida, as posturas, ou guardas, são as presentes na Tabela 2.1.

Figura	Nome da guarda	Descrição
	<i>Ox [Ochs]</i>	À direita: posicione seu pé esquerdo à frente e segure a espada com a empunhadura próxima ao rosto apontada para frente. À esquerda: faça o mesmo mas com o pé direito à frente.
	<i>Plow [Pflug]</i>	À direita: posicione seu pé direito à frente e segure a espada com a empunhadura próxima ao seu joelho da frente, apontada pra frente. À esquerda: faça o mesmo mas com o pé esquerdo à frente.
	<i>Day [Tag]</i>	Posicione seu pé esquerdo à frente e segure a espada com a ponta para cima.
	<i>Fool [Olber]</i>	Posicione seu pé esquerdo à frente e segure a espada com a ponta para baixo, em frente ao seu pé direito.

Tabela 2.1: As 4 guardas principais segundo Meyer (1570). Figuras disponíveis em Wiktenauer (2020).

Além das guardas, é importante dar atenção à base, ou seja, à posição das pernas, de forma a manter o equilíbrio e facilitar a movimentação. Um conceito fácil de assimilar para o entendimento de uma boa postura é o conceito da linha central. A linha central é uma linha imaginária, usada como referência em diversas artes marciais, que passa pelo centro de massa do praticante e aponta para frente. Idealmente, o praticante deve manter o pé e o joelho da frente apontados nessa direção, assim como o rosto e o tronco.

Uma forma simples de avaliar se a postura (isolando apenas a parte das pernas) está adequada é garantir que o pé da frente esteja paralelo à linha central e o pé de trás não a atravesse. O pé de trás também deve estar posicionado ligeiramente para fora diagonalmente, pois, deste modo, o praticante mantém seu centro de massa no meio de sua base e seus calcanhares são capazes de distribuir forças frontais e laterais para suas pernas, evitando um desequilíbrio. Isso pode ser visualizado mais facilmente com auxílio da Tabela 2.2.

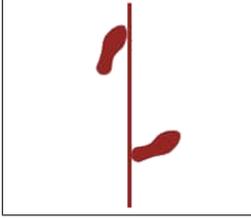
Posição dos pés	Descrição
	Ideal.
	Pé de trás atravessa a linha central, o praticante perde estabilidade.
	Aceitável, mas os pés estão distantes horizontalmente.
	Ambos os pés estão apontados para frente, o praticante perde estabilidade.
	Aceitável, mas os pés estão muito juntos. Ideal para outras artes marciais, como o boxe.

Tabela 2.2: Posições dos pés em relação à linha central (BoxingIsOn, 2021).

3 Desenvolvimento

Neste capítulo são apresentados os detalhes de implementação da ferramenta para verificação das posturas discutidas no capítulo anterior. O programa foi inicialmente planejado para funcionar da seguinte forma:

1. Recebe uma imagem ou um conjunto de imagens do usuário;
2. Usa as bibliotecas *OpenPose* e *TensorFlow* para adquirir os modelos de esqueleto, mãos e objeto;
3. Calcula qual postura da base de dados é a mais próxima da postura identificada;
 - (a) Primeiro cálculo apenas com os pontos de articulação da cintura para baixo, pois a posição das pernas é comum a todas as guardas básicas;
 - (b) Segundo cálculo com os pontos de articulação da cintura para cima, que diferem entre as guardas.
4. Calcula a variação entre as duas posturas;
5. Calcula se os pés estão alinhados corretamente, independentemente da postura;
6. Informa ao usuário o quão próxima a sua postura está do ideal, com base em *thresholds* adquiridos empiricamente a partir dos dados de instrutores certificados da arte marcial. A informação é separada membro a membro, de forma que o usuário entenda exatamente como corrigir sua postura.

Os dados inicialmente disponíveis são os modelos de esqueleto e de mãos e a posição do objeto (na forma (x_1, y_1, x_2, y_2) , como um retângulo), ou seja, as posições dos pontos de articulação no espaço, sendo que todos são numerados. A partir desses pontos, podemos calcular os vetores ou arestas que representam os membros, permitindo computar as distâncias entre os pontos de articulação, as posições dos pontos de articulação relativas a outros pontos de articulação, os ângulos dos membros relativos ao eixo cartesiano e os ângulos dos membros relativos a outros membros.

Os valores de distância podem ser prontamente descartados pois são irrelevantes neste contexto. Diferentes pessoas possuem diferentes proporções corporais, além de o ângulo e a distância da câmera poderem interferir na inferência da postura. A própria resolução da imagem causaria problemas se as distâncias não fossem normalizadas primeiro. Além disso, comparados aos outros dados, os valores de distância não apresentam um ganho significativo na inferência.

Os ângulos relativos ao eixo cartesiano também não são ideais, pois os relativos a membros adjacentes são mais precisos. Apenas a coluna pode ser medida em relação ao eixo cartesiano para melhores resultados, em vez de em relação à linha do quadril. Os ângulos entre membros adjacentes foram também os que apresentaram menor variação para posturas iguais fotografadas de ângulos diferentes.

A posição de um ponto em relação a outro (esquerda, direita, cima, baixo) ajuda a máquina a tratar certas ambiguidades. Por exemplo, como a máquina lida majoritariamente com ângulos, ela pode ter dificuldade de tratar certas exceções, como pode ser visto na Figura 3.1. Nela, há duas fotos tiradas do mesmo ângulo, uma na guarda *Alber* e a outra na guarda *Vom Tag*, ambas com a perna esquerda à frente. Mas, devido à diferença na rotação do tronco, o ângulo entre a perna e a linha do quadril é completamente diferente. Em uma imagem as pernas aparentam estar cruzadas perto do quadril, mas é perceptível que elas não estão de fato cruzadas. Trata-se de uma impressão causada pelo ângulo em que a foto foi tirada.

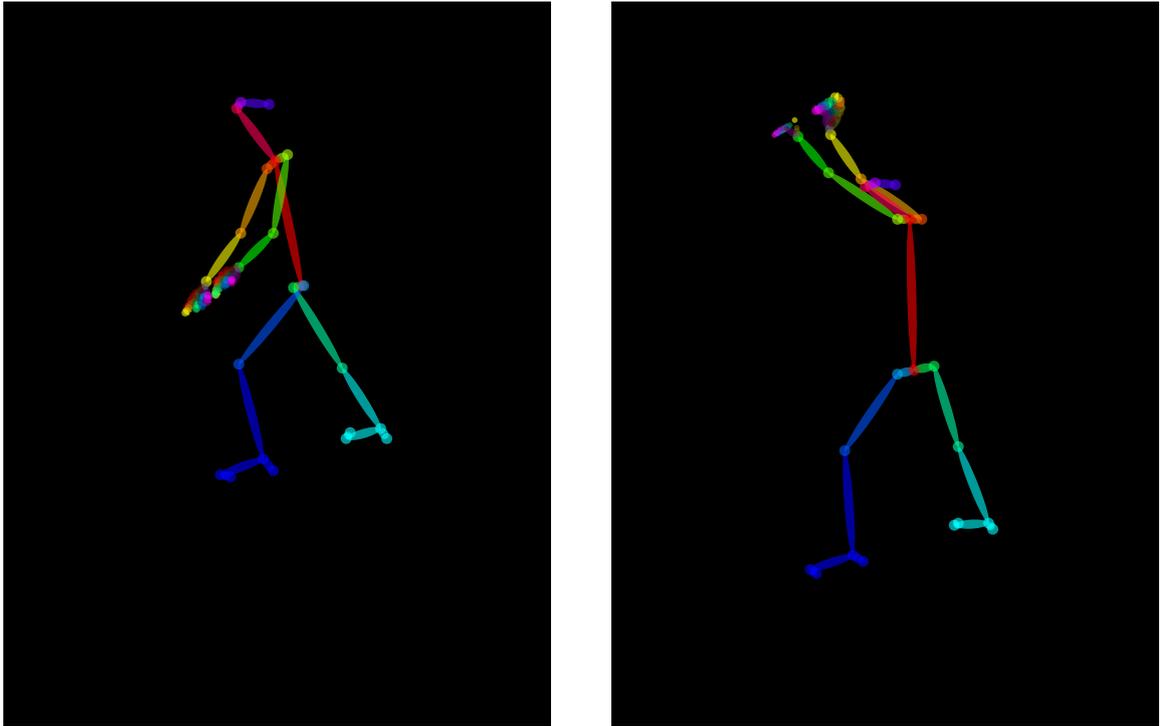
(a) *Alber*, pernas “cruzadas”.(b) *Vom Tag*, pernas “descruzadas”.

Figura 3.1: Desambiguação da postura de pernas.

A base de dados é construída a partir de imagens como essas, mas são retirados apenas os valores mais relevantes, de modo que diversos modelos diferentes das pernas e dos braços usados como referência pelo programa são armazenados em arquivos *.csv* separados (um arquivo para pernas e outro para braços, sendo que a inclinação da coluna fica armazenada no arquivo das pernas por conveniência). O arquivo é composto por uma única tabela onde cada linha representa uma postura diferente capturada de um ângulo diferente e cada coluna representa um ângulo diferente entre os pontos de articulação do modelo.

Os dados relevantes extraídos são então comparados entre todas as 4 guardas (6 se contarmos as variações esquerda/direita de *Ochs* e *Pflug*), cada uma fotografada de 5 ângulos diferentes: frontal, lateral à esquerda, lateral à direita, diagonal à esquerda e diagonal à direita, constituindo um total de 30 modelos corretos diferentes. Além dos modelos corretos, foram fotografadas posturas completamente diferentes, visto que alguns erros são sutis para a ferramenta e podem passar despercebidos pelo cálculo de aproximação. A solução adotada foi criar novas linhas para a tabela com diferentes posturas erradas, fazendo a ferramenta reconhecer posturas erradas específicas, similares a postu-

ras corretas. Posturas erradas foram pensadas com base na Tabela 3.1. Ao embaralhar o conteúdo da tabela, são adquiridas posturas erradas que poderiam ser aproximadas das corretas pela ferramenta. Além destas, algumas outras posturas erradas foram adicionadas à base de dados com base em ambiguidades detectadas durante a implementação, com erros sutis para o olho humano mas com números completamente distoantes para a máquina, como pernas cruzadas e ângulos desfavoráveis.



Figura 3.2: Os 5 ângulos de captura. Imagens extraídas de Rüter (2016).

3.1 As posturas

Com o intuito de classificar facilmente as diferentes posturas, algumas características devem ser destacadas. Primeiramente, as pernas estão sempre posicionadas da mesma forma:

- A perna da frente deve estar flexionada;
- A perna de trás deve estar estendida;
- O pé da frente deve apontar para frente, paralelo à linha central;
- O pé de trás deve apontar diagonalmente para fora da linha central.

Com base nas descrições de Meyer, pode ser feita a primeira abstração sobre as guardas e a forma como elas podem ser classificadas:

3.2 Orientação dos pés e calcanhares

Como visto no capítulo anterior, a posição dos pés está correta se o pé da frente estiver apontando para frente em relação à linha central e se o pé de trás não a atravessar. Para

Postura	Perna da frente	Posição das mãos	Sentido da espada
<i>Vom Tag</i>	Esquerda	Acima da cabeça	Para cima
<i>Alber</i>	Esquerda	Apontadas para baixo	Para baixo
<i>Pflug</i> à esquerda	Esquerda	Próximas ao joelho da frente	Para frente
<i>Pflug</i> à direita	Direita	Próximas ao joelho da frente	Para frente
<i>Ochs</i> à esquerda	Direita	Na altura dos olhos	Para frente
<i>Ochs</i> à direita	Esquerda	Na altura dos olhos	Para frente

Tabela 3.1: Primeira abstração das posturas.

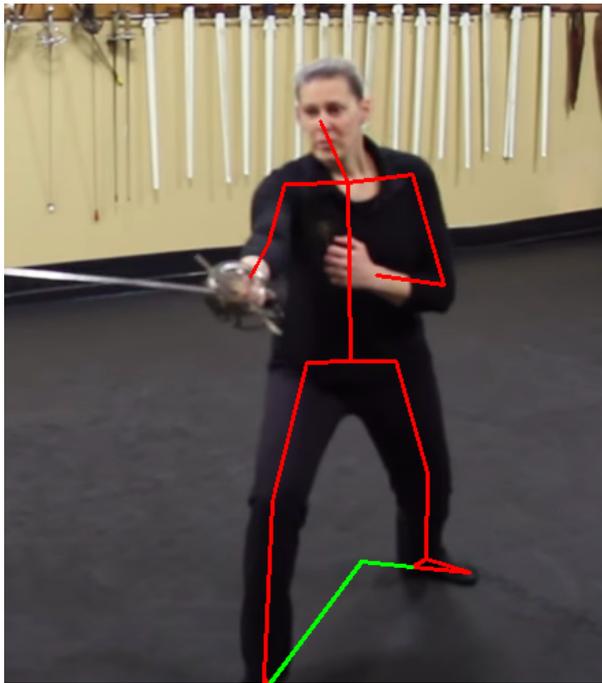
esse cálculo, avaliam-se os dados dos calcanhares do modelo BODY_25 (já que o modelo COCO possui menos pontos de articulação para os pés). Devem ser traçadas duas linhas paralelas ao sentido de cada pé. Para a postura ser considerada correta, a interseção entre essas linhas deve existir e deve ocorrer atrás de ambos os tornozelos. Considerando os pontos 19, 21, 22 e 24 do modelo, foi escrita a função que pode ser vista na Listagem 3.1:

Listagem 3.1: Trecho do código onde é calculada a interseção das linhas dos pés.

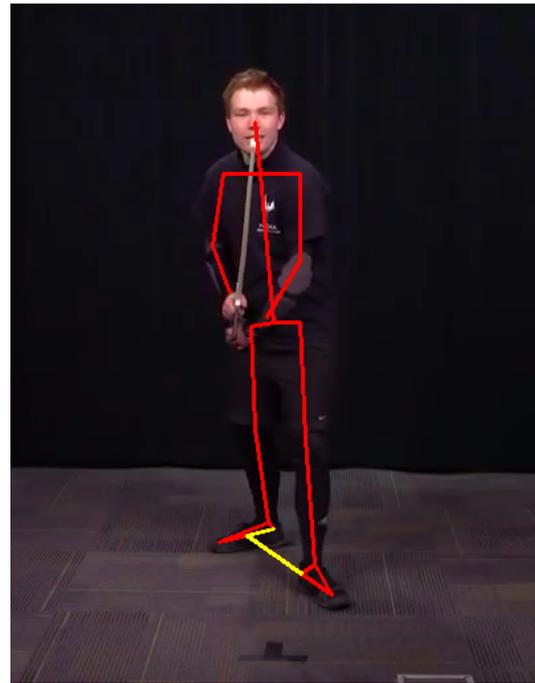
```

1  const auto a = datumsPtr->at(0)->poseKeypoints[{0, 19, 0}];
2  const auto b = datumsPtr->at(0)->poseKeypoints[{0, 19, 1}];
3  const auto c = datumsPtr->at(0)->poseKeypoints[{0, 21, 0}];
4  const auto d = datumsPtr->at(0)->poseKeypoints[{0, 21, 1}];
5  const auto e = datumsPtr->at(0)->poseKeypoints[{0, 22, 0}];
6  const auto f = datumsPtr->at(0)->poseKeypoints[{0, 22, 1}];
7  const auto g = datumsPtr->at(0)->poseKeypoints[{0, 24, 0}];
8  const auto h = datumsPtr->at(0)->poseKeypoints[{0, 24, 1}];
9
10 // encontra o ponto (x) de intersecao usando determinantes
11 double d1 = a * d - b * c;
12 double d2 = e * h - f * g;
13 double d3 = (a - c) * (f - h) - (b - d) * (e - g);
14 double x  = ((d1 * (e - g)) - (d2 * (a - c))) / d3;
15
16 // intersecao esta a frente do calcanhar?
17 return ((a > c && x > c) ||
18         (a < c && x < c) ||
19         (e > g && x > g) ||
20         (e < g && x < g));

```



(a) Interseção atrás dos calcanhares (Smith et al, 2017).



(b) Interseção à frente dos calcanhares (Allen, 2016).

Figura 3.3: Interseção das linhas dos pés para avaliação da corretude da base (módulo de detecção de objetos desligado).

3.3 Detecção da arma

Com todos os dados relevantes do corpo do praticante adquiridos, resta calcular a posição da arma no espaço. Para isso, 3 métodos foram considerados:

- Usar os pontos de articulação dos pulsos: método mais impreciso, mas funcional. Pode ser usado para imagens de baixa resolução ou quando por algum outro motivo a máquina não é capaz de encontrar as mãos do praticante na imagem;
- Calcular a rotação da arma como a média das regressões lineares dos pontos das 2 mãos: apresenta resultados melhores que os pontos de articulação dos pulsos, mas não por uma margem muito significativa;
- Usar o módulo de detecção de objetos do *TensorFlow* para encontrar a arma no espaço: método mais preciso, mas nem sempre a RNA detecta o objeto na imagem. Portanto, ainda é necessário usar outro método para esse caso. Como o módulo de detecção de objetos adquire apenas o retângulo onde o objeto se encontra na

imagem, é necessário traçar a reta entre o centro do retângulo e a mão que segura a arma (se a arma estiver empunhada em duas mãos, pode ser usada qualquer uma das mãos).

3.4 Correção da inferência

Como a ferramenta pode apresentar erros na inferência da postura do praticante, é realizada uma análise do conjunto de dados para corrigir ou mitigar esse erro. Mais especificamente, como a base (postura das pernas) é sempre a mesma independente da guarda (trocando apenas a perna da frente), a diferença é apenas o ângulo em que ela foi fotografada. É comparada então a inferência do ângulo da câmera com o ângulo de captura da guarda detectada, tendo em vista que, assim como a base, cada guarda possui variações na base de dados de acordo com o ângulo em que foi fotografada. Os valores possíveis para ambos são: frontal, lateral (esquerda e direita) e diagonal (esquerda e direita).

A depender da diferença entre elas, as seguintes ações devem ser tomadas:

- Idênticas: nenhuma ação;
- Esquerda *vs* direita: prevalece o ângulo da câmera em detrimento da base;
- Lateral *vs* diagonal: o modelo adquirido é comparado ao mesmo tempo com os modelos lateral e diagonal da base de dados para a guarda em questão, e é escolhido sempre o menor valor de diferença, aumentando assim a leniência da ferramenta;
- Frontal *vs* não frontal: assume-se que a base está completamente errada e prevalece o ângulo da câmera.

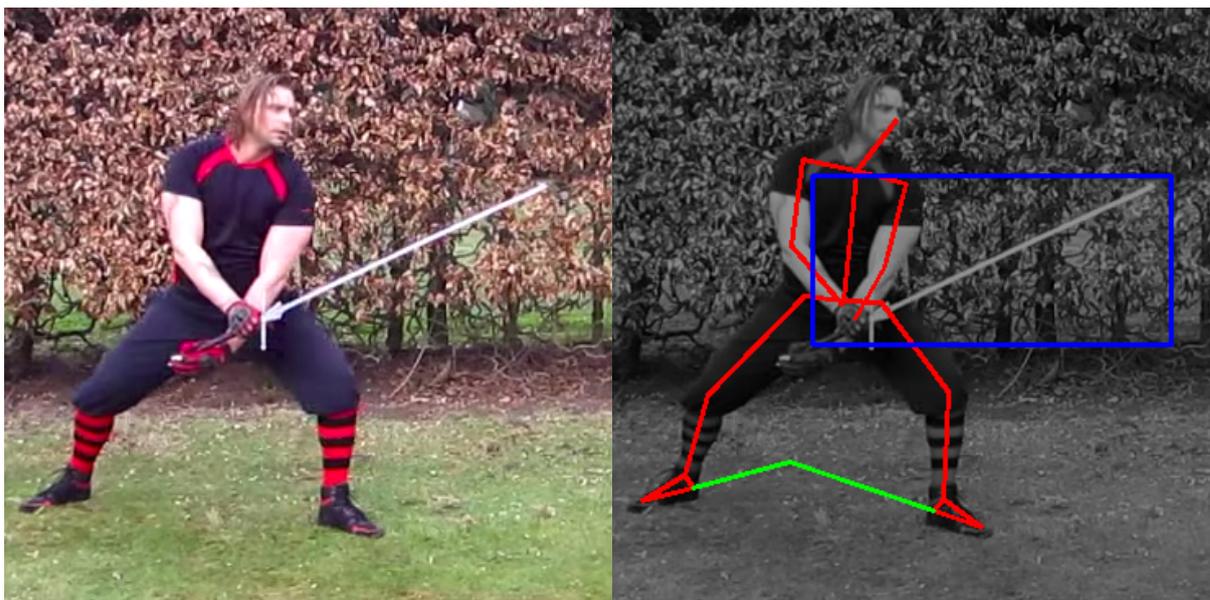
É importante ressaltar que a guarda em si não pode ser corrigida. O que pode ser corrigido é o ângulo da câmera e o ângulo de captura da base (da cintura para baixo).

4 Análise de dados

Foram usadas 50 imagens para criar a base de dados. Desse conjunto, 30 imagens são de cada uma das 4 guardas capturadas de 5 ângulos diferentes (lateral esquerdo, diagonal esquerdo, frontal, diagonal direito e lateral direito), lembrando que 2 dessas 4 guardas possuem variação lateral, resultando em um total de 6 guardas. Há ainda uma imagem que é especificamente a guarda *Ochs* à direita capturada do ângulo lateral esquerdo, onde a cabeça oculta o braço esquerdo na imagem. Isso é comum para essa guarda quando fotografada desse ângulo, como pode ser visto na Figura 3.2. Por fim, 19 imagens são de possíveis variações de imagens das pernas capturadas de ângulos diferentes, com ou sem cruzamento das pernas, e com a perna esquerda ou a direita na frente.

Para testar a ferramenta, foram usadas 150 imagens, sendo elas 24 da guarda *Alber*, 23 de *Ochs* à esquerda, 30 de *Ochs* à direita, 27 de *Pflug* à esquerda, 17 de *Pflug* à direita e 29 de *Tag*. Cada subgrupo de imagens possui ao menos imagens capturadas de cada um dos 5 ângulos possíveis e de pelo menos 3 praticantes diferentes. Dentre essas 150 imagens, há posturas próximas às corretas e também posturas muito distantes, difíceis de serem detectadas. Adicionalmente, foram usadas outras 44 imagens juntamente com as 50 da base de dados, para definir os *thresholds*. Entretanto, essas imagens foram descartadas após a consolidação da base de dados.

As imagens foram adquiridas com uma câmera de celular comum ou tiradas de quadros de vídeos de instrutores. Para cada imagem é feita a verificação da guarda, da base e do ângulo da câmera. As guardas são as 4 mencionadas anteriormente: *Alber*, *Ochs* (esquerda e direita), *Pflug* (esquerda e direita) e *Vom Tag*. Também é verificada a perna da frente e se os calcanhares estão na posição correta, mas a taxa de acerto para ambos não é de 100% apenas porque em algumas imagens não é possível adquirir o modelo correto da pessoa devido a alguma sobreposição na foto. Por exemplo, devido ao ângulo da câmera em relação à pessoa, uma perna pode estar atrás da outra, interferindo na aquisição do modelo com o *OpenPose*. As imagens da Figura 4.1 ilustram como a ferramenta funciona.



(a) Imagem original (Rüther, 2016).

(b) Imagem processada.

Figura 4.1: Funcionamento da ferramenta.

Ao final do processamento, são exibidos no *console* os valores dos ângulos entre as articulações marcadas na segunda imagem da Figura 4.1. Tudo que importa para a ferramenta são esses ângulos, a interseção da linha dos calcanhares e a posição da espada. Todo o resto é descartado.

A ferramenta foi executada numa máquina com 16GB de memória RAM, placa gráfica Nvidia GTX 1050 Ti e processador Intel Core i7-8750H 2,20GHz de 12 núcleos. A inicialização dos módulos das bibliotecas *OpenPose* e *TensorFlow* leva em média 1,5s e cada imagem é processada em aproximadamente 0,5s. O menor tempo registrado foi de 0,39s para uma imagem de 400x400 *pixels*, e imagens de resolução 1920x1080 (maior resolução usada nos testes) podem subir o tempo de processamento individual para até 0,68s.

A precisão para as classificações das 150 imagens de teste pode ser vista no gráfico da Figura 4.2:

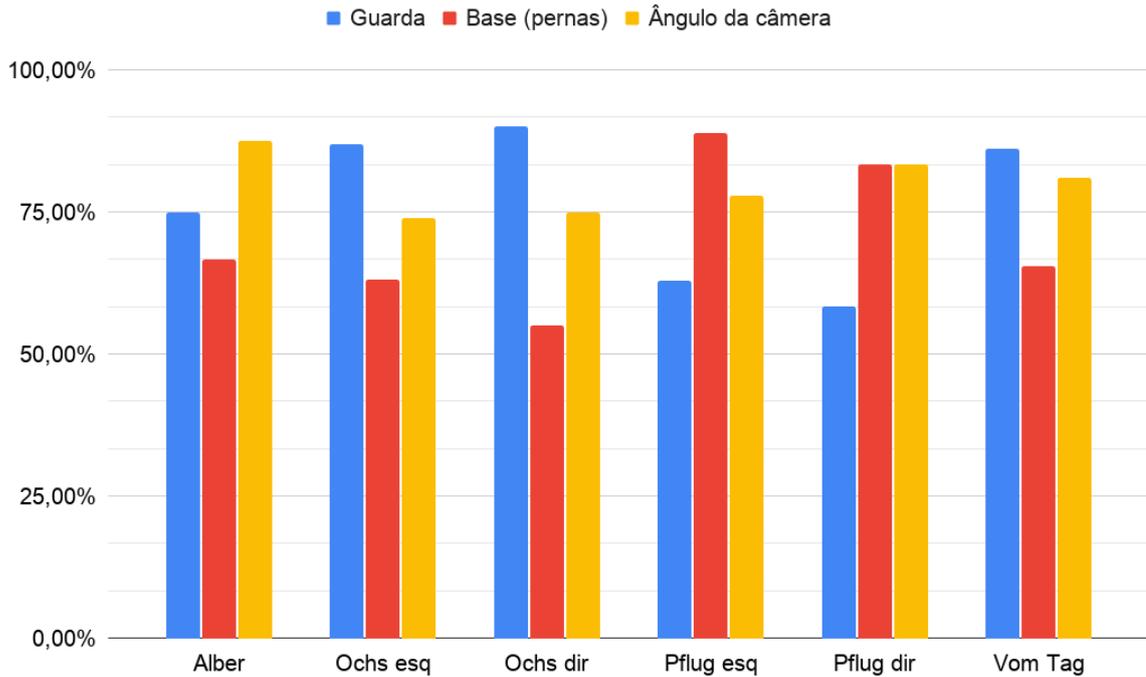


Figura 4.2: Taxa de acerto da ferramenta sem a detecção de objetos e sem regressão linear.

Os resultados iniciais indicam uma taxa de acerto próxima de 90% em alguns casos. Entretanto, temos outras taxas que ficam pouco acima de 50%. Vale lembrar que, apesar de mostrados separadamente no gráfico, os resultados são analisados em conjunto, portanto mesmo quando a ferramenta não acerta todas as informações, ainda é possível corrigir essa inferência, ou ao menos mitigar o erro. Isso não aumenta a precisão máxima da ferramenta, mas reduz o erro para os valores mais imprecisos, subindo a precisão mínima consideravelmente.

O que prevalece para essa correção é a guarda detectada. O ângulo da câmera pode ser corrigido, mas a guarda não. As guardas *Alber* e *Pflug* são muito similares para a ferramenta, o que dificulta a detecção de ambas. No gráfico da Figura 4.2, é imediato notar que a taxa de acerto na detecção das guardas baixas (*Alber* e *Pflug*) não é alta. Mas o módulo de detecção de objetos do *TensorFlow* aumenta a precisão da inferência da guarda por uma margem significativa, como mostra a Figura 4.3.

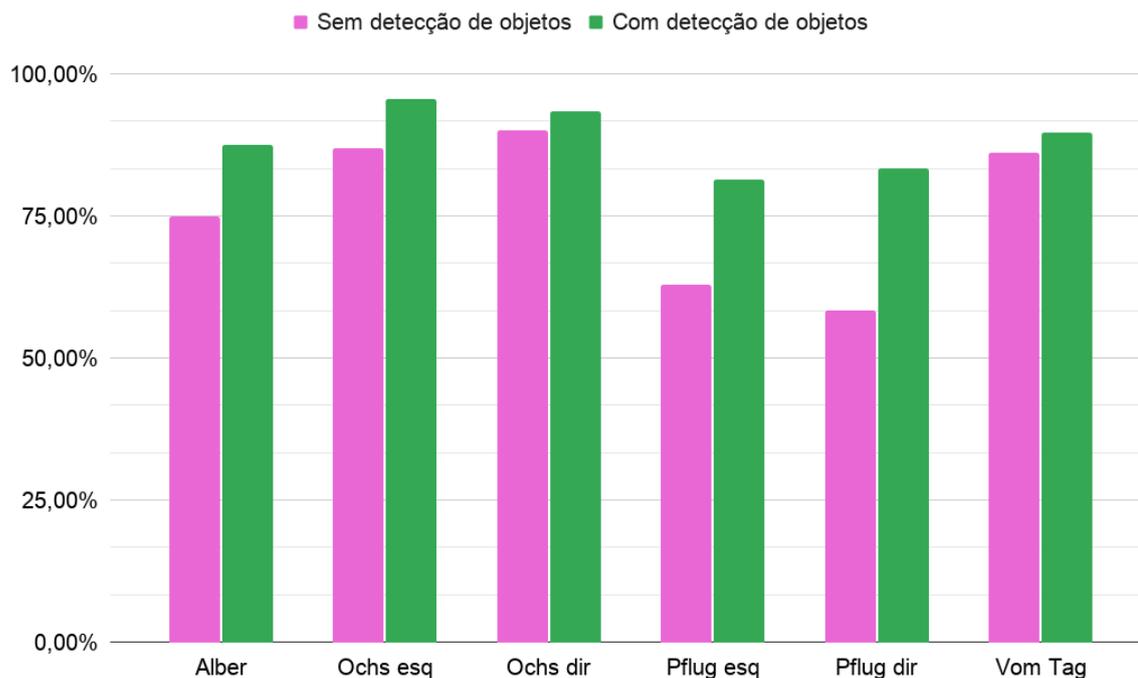


Figura 4.3: Taxa de acerto com e sem detecção de objetos (e sem regressão linear).

Como pode ser observado no gráfico, a taxa de acerto para a guarda *Alber* subiu de 75% para 87,5%, e a da guarda *Pflug* subiu de aproximadamente 60% para 83%. O módulo de detecção de objetos não só aumenta significativamente a precisão de inferências anteriormente imprecisas, mas também aumenta ligeiramente a precisão de inferências já satisfatórias. Isso ocorre porque as posturas baixas (*Alber* e *Pflug*), quando há qualquer ligeira imprecisão na detecção das mãos, são facilmente confundidas, pois o valor mais contrastante entre essas posturas é a direção da espada, como pode ser visto na Tabela 3.1.

Outro fator que contribui para esse aumento de precisão é que os pontos das mãos, quando fotografados frontalmente, ficam extremamente próximos, produzindo assim uma enorme diferença entre o ângulo esperado e o ângulo detectado pela ferramenta. Já com a detecção de objetos, esse ângulo é medido entre dois pontos muito mais distantes na foto (as mãos e o centro do retângulo que contém a espada), diminuindo significativamente essa imprecisão.

Comparativamente, o outro método discutido no capítulo anterior, de detectar a postura relativa dos membros superiores e a inclinação da arma no espaço através dos pontos de articulação da mão do praticante, apresentam resultados menos satisfatórios,

como pode ser visto no gráfico da Figura 4.4. Isso acontece porque em alguns casos os pontos da mão ficam muito próximos, ou pode haver uma leve oclusão que faz a ferramenta errar a localização dos pontos por uma distância pequena na imagem. Então traçar uma reta, seja com uma regressão linear ou com apenas dois pontos, pode resultar em um ângulo totalmente inesperado. Na Figura 4.1, por exemplo, há uma pequena oclusão da mão direita sobre o braço esquerdo. Isso basta para que a linha dos pulsos (detectada pela ferramenta) aponte para baixo. Então, é seguro assumir que a postura *Pflug* da imagem seria confundida com *Alber*, pois a maior diferença entre elas é o sentido da espada. Entretanto, a regressão linear ainda apresenta, em média, resultados melhores que traçar a reta a partir dos pulsos:

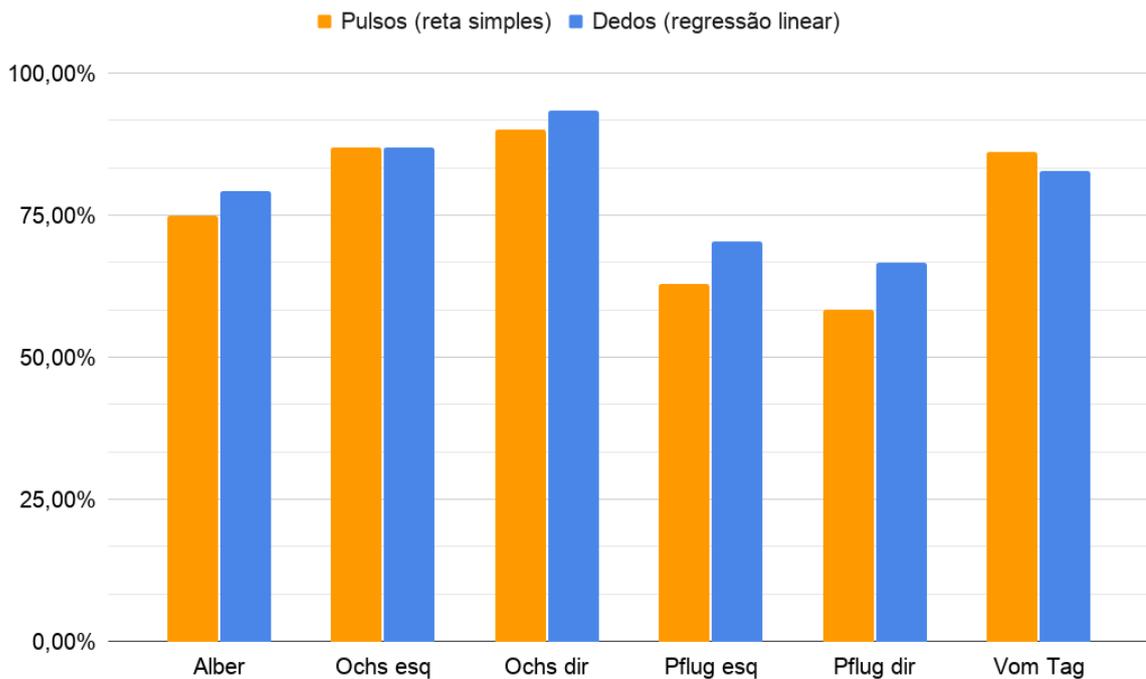


Figura 4.4: Taxa de acerto com e sem regressão linear (sem detecção de objetos).

Conclui-se, portanto, que o módulo de detecção de objetos é essencial para a ferramenta, pois sem ele não haveria solução simples para os problemas de imprecisão encontrados.

4.1 Limitação da base de dados

Devido à pandemia de 2020, não foi possível realizar testes extensivos com diversos praticantes como inicialmente planejado. Então os resultados podem ser considerados tendenciosos ou enviesados por terem sido realizados em um universo estatístico muito pequeno. Entretanto, é possível que, com um conjunto de dados maior, a ferramenta receba um conjunto de dados de treinamento maior e, conseqüentemente, adquira uma precisão maior no geral.

Independentemente disso, com o algoritmo de treinamento já implementado, é trivial passar uma nova base de dados para a ferramenta, sendo necessários, possivelmente, apenas alguns ajustes nos valores de *threshold*.

5 Conclusão

Neste trabalho foi apresentada a proposta de avaliar, com base em imagens 2D e com auxílio das ferramentas *OpenPose* e *TensorFlow*, posturas descritas no livro de Forgen (2015), de forma a criar uma ferramenta para o treino autodidata de HEMA. Foi criada uma base de dados com os ângulos entre as articulações mais importantes, gerada a partir de fotografias dessas posturas, e implementado um algoritmo para a comparação de novas posturas, ajustando valores de *threshold* para delimitar o quanto uma postura pode ser diferente da existente na base de dados e ainda ser considerada correta no contexto da arte marcial.

Levando em consideração tudo o que foi visto até este ponto, pode-se concluir que os resultados são satisfatórios o bastante para criar uma aplicação prática, mas apenas para ser usada de forma supervisionada, com um instrutor ou praticante mais experiente ao lado. A taxa de acerto média de 88,49% da ferramenta com o módulo de detecção de objetos é boa mas, para as posturas com menor taxa de acerto (aproximadamente 80%), um falso positivo poderia ser interpretado de maneira errada por iniciantes com uma frequência preocupante. Acredita-se que é possível atingir números melhores com o aprimoramento da ferramenta e com uma base de dados mais robusta. Entretanto, os resultados apresentados já podem ser usados para demonstrar quais ferramentas e técnicas são mais adequadas para trabalhos similares.

Este trabalho foi desenvolvido especificamente para a prática de HEMA, mas é simples adaptar a base de dados existente para qualquer arte marcial, com ou sem armas brancas. Cada arte marcial possui posturas diferentes, então cada uma deve possuir sua própria base de dados, mas é possível usar o mesmo programa, ou mesmo criar um *framework* para várias artes marciais diferentes. Isso foge do escopo deste trabalho, mas é uma tarefa factível para pesquisadores interessados. Essa tarefa se torna ainda mais simples se considerarmos que as posturas das pernas são comuns à maioria das artes marciais. Dessa forma, seria necessário adaptar somente a base de dados dos braços e das mãos. Para a maioria das artes marciais, inclusive, o módulo de detecção de objetos é

desnecessário, uma vez que não fazem uso de armas brancas.

Outra possibilidade para trabalhos futuros é trabalhar com movimentos da arte marcial, e não apenas com posturas estáticas. Para isso seria necessário um conjunto de imagens muito maior e uma ferramenta que seja capaz de computar as imagens em tempo real. Para contornar essa dificuldade, entretanto, pode ser usado um módulo de reconstrução tridimensional ou até uma câmera com sensor ToF (*Time of Flight*), também chamado “sensor de profundidade”, que é capaz de determinar a distância de objetos capturados pela câmera. Exemplos de dispositivos com esse sensor são o *Kinect* e modelos mais recentes de *smartphones*. Dessa forma, seria possível abranger uma gama muito maior de técnicas da arte marcial, visto que toda arte marcial é uma prática dinâmica.

Bibliografia

- Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y. ; Zheng, X. **TensorFlow: Large-scale machine learning on heterogeneous systems**, 2015. Software available from tensorflow.org.
- Allen, N. **HEMA at VCU**, 2016 (acessado em: 18 mar. 2021). Disponível em: https://www.youtube.com/channel/UCuiZKL4W4yFU12yv9m_H4LA.
- Bansal, A. Performance analysis of sign language detection using deep neural networks and computer vision. **group**, v.5, p. 6, 2020.
- Barrantes, S.; Egea, A. J. S.; Rojas, H. A. G.; Martí, M. J.; Compta, Y.; Valldeoriola, F.; Mezquita, E. S.; Tolosa, E. ; Valls-Solè, J. Differential diagnosis between parkinson's disease and essential tremor using the smartphone's accelerometer. **PloS one**, v.12, n.8, p. e0183843, 2017.
- BoxingIsOn. **Boxing workout for beginners (BoxingIsOn)**, Acessado em: 10 fev. 2021. Disponível em: <https://www.boxingison.com/boxing-workouts-for-beginners/>.
- Cao, Z.; Hidalgo Martinez, G.; 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.
- Cao, Z.; Hidalgo Martinez, G.; Simon, T.; Wei, S. ; Sheikh, Y. A. **OpenPose (GitHub)**, 2019 (acessado em: 1 fev. 2021). Disponível em: <https://github.com/CMU-Perceptual-Computing-Lab/openpose>.
- Chen, Y.; Shen, C. Performance analysis of smartphone-sensor behavior for human activity recognition. **Ieee Access**, v.5, p. 3095–3110, 2017.
- 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, 2020.
- Cynarski, W. J.; Wajs, W. ; Vences Brito, A. Improving the movements of basic karate techniques with the use of motion capture and mathematical modeling. the concept of a research project. **Ido Movement for Culture. Journal of Martial Arts Anthropology**, v.1, n.14, p. 39–53, 2014.
- Edward, P. **Golf club accelerometer**, Dez. 28 1965. US Patent 3,226,704.
- Evans, J. W. **Athletic swing measurement system**, Set. 6 1966. US Patent 3,270,564.

- de Farias, T. **Metodologia para reconstrução 3d baseada em imagens**. 2012. Tese de Doutorado - Ph. D. dissertation, Centro de Informática, Universidade Federal de . . .
- Forgeng, J. L. **The Art of Combat**. Frontline Books, 2015.
- Haykin, S. **Redes neurais: princípios e prática**. Bookman Editora, 2007.
- He, T.; Luo, Q. **A survey of motion capture technology and its application in sports**. In: International Conference on Human Systems Engineering and Design: Future Trends and Applications, p. 854–859. Springer, 2018.
- Kamaya, Y.; Naka, T.; Yamada, M. ; Miyazaki, S. **Information visualization for virtual martial arts training**. In: 2018 Nicograph International (NicoInt), p. 66–69. IEEE, 2018.
- Kovács, Z. L. **Redes neurais artificiais**. Editora Livraria da Fisica, 2002.
- Merkert, H. **A biomechanical analysis of technique training and freefight in historical european martial arts**, 2019.
- Meyer, J. **Gründtliche Beschreibung der Kunst des Fechtens**. Strassburg: printed by Thiebolt Berger, 1570.
- de Milano, D.; Honorato, L. B. **Visão computacional**, 2014.
- Moeslund, T. B.; Granum, E. A survey of computer vision-based human motion capture. **Computer vision and image understanding**, v.81, n.3, p. 231–268, 2001.
- Nicholas, V. **Golf drive metering apparatus**, Mai 11 1965. US Patent 3,182,508.
- Polak, E.; Kulasa, J.; Vences Brito, A.; Castro, M. A. ; Fernandes, O. Motion analysis systems as optimization training tools in combat sports and martial arts. **Revista de Artes Marciais Asiáticas**, v.10, n.2, p. 105–123, 2016.
- Reyes-Ortiz, J.-L.; Oneto, L.; Samà, A.; Parra, X. ; Anguita, D. Transition-aware human activity recognition using smartphones. **Neurocomputing**, v.171, p. 754–767, 2016.
- Robert-Lachaine, X.; Mecheri, H.; Muller, A.; Larue, C. ; Plamondon, A. Validation of a low-cost inertial motion capture system for whole-body motion analysis. **Journal of biomechanics**, v.99, p. 109520, 2020.
- Rüther, B. **Björn Rüther**, 2016 (acessado em: 27 fev. 2021). Disponível em: https://www.youtube.com/watch?v=v2Sh7N6c1HM&ab_channel=Bj%C3%B6rnR%C3%BCther.
- Shoaib, M.; Bosch, S.; Incel, O.; Scholten, H. ; Havinga, P. Fusion of smartphone motion sensors for physical activity recognition. **Sensors**, v.14, n.6, p. 10146–10176, 2014.
- Shoaib, M.; Bosch, S.; Incel, O.; Scholten, H. ; Havinga, P. Complex human activity recognition using smartphone and wrist-worn motion sensors. **Sensors**, v.16, n.4, p. 426, 2016.
- Franklin, S.; Smith, N. **Blood and Iron HEMA**, 2017 (acessado em: 14 dez. 2020). Disponível em: https://www.youtube.com/watch?v=00gaIuzR0rs&t=173s&ab_channel=BloodandIronHEMA.

- Vences Brito, A. M.; Ferreira, M. A. R.; Cortes, N.; Fernandes, O. ; Pezarat-Correia, P. Kinematic and electromyographic analyses of a karate punch. **Journal of Electromyography and Kinesiology**, v.21, n.6, p. 1023–1029, 2011.
- White, W. P. **Speed-indicating device for golf clubs**, Mai 14 1929. US Patent 1,712,537.
- Wiktenauer. **Joachim Meyer (Wiktenauer)**, 2020 (acessado em: 13 dez. 2020). Disponível em: https://wiktenauer.com/index.php?title=Joachim_Meyer&oldid=120195.
- Worsey, M. T.; Espinosa, H. G.; Shepherd, J. B. ; Thiel, D. V. Inertial sensors for performance analysis in combat sports: a systematic review. **Sports**, v.7, n.1, p. 28, 2019.
- Wu, W.; Dasgupta, S.; Ramirez, E. E.; Peterson, C. ; Norman, G. J. Classification accuracies of physical activities using smartphone motion sensors. **Journal of medical Internet research**, v.14, n.5, p. e130, 2012.
- Zhang, W.; Liu, Z.; Zhou, L.; Leung, H. ; Chan, A. B. Martial arts, dancing and sports dataset: A challenging stereo and multi-view dataset for 3d human pose estimation. **Image and Vision Computing**, v.61, p. 22–39, 2017.