Sistema de Classificação de Sinalética Gestual em Competições de Karaté
Gesture Classification System for Karate Competitions

Sónia Correia Violante
Department of Science and Technology, Universidade Aberta, 4200-055 Porto, Portugal
1100135@estudante.uab.pt

A. Jorge Morais
Department of Science and Technology, Universidade Aberta, 4200-055 Porto, Portugal
INESC TEC–INESC Tecnologia e Ciência, 4200-465 Porto, Portugal
LE@D, Laboratory of Distance Education and eLearning, Universidade Aberta, 1269-001 Lisbon, Portugal
jorge.morais@uab.pt

Vítor Filipe
School of Science and Technology, University of Trás-os-Montes e Alto Douro, Vila Real, Portugal
INESC TEC–INESC Tecnologia e Ciência, 4200-465 Porto, Portugal
vfilipe@utad.pt

Resumo — **Em contexto de Kumite (combate de Karate) propõe-se investigar um modelo de classificação da sinalética gestual do árbitro para atribuição de pontos, com recurso a *Visão Computacional* e técnicas de *Aprendizagem Profunda*. Foram realizadas três abordagens, todas tendo como base o recurso a modelos de Redes Neuronais Convolucionais (Convolutional Neural Network – CNN): Classificação de imagens com recurso a uma CNN; Deteção da pose humana com o modelo *MoveNet*; e a deteção e classificação de gestos com o modelo *YOLOv5*, via *RoboFlow*. A última abordagem obteve melhores resultados, com 100% de *precision* para todas as classes, pelo que se testou a sua aplicação para a deteção e classificação dos gestos em vídeo.**

Palavras Chave – Visão Computacional; Aprendizagem Profunda; Redes Neuronais Convolucionais; Karate; Sinalética gestual; Árbitro; RoboFlow.

Abstract — **In the context of Kumite (Karate combat), it is proposed to investigate a model for classifying the referee's gestures to award points using Computer Vision and Deep Learning techniques. Three approaches were used, all based on Convolutional Neural Networks (CNN) models: image classification using a CNN; human pose detection using *MoveNet*; and object detection using *YOLOv5*, via *RoboFlow*. The last approach obtained the best results, with 100% precision for all classes, so we tested its application for video gesture detection and classification.**

Keywords - Computer Vision; Deep Learning; Convolutional Neural Network; Karate; Gesture; Referee; RoboFlow.

I. Introdução

A Visão Computacional, como campo interdisciplinar da Inteligência Artificial, em conjunto com técnicas de Aprendizagem Profunda, permite que os sistemas sejam capazes de interpretar e entender informações visuais a partir de imagens ou vídeos. Por outras palavras, este tipo de sistema é capaz de compreender o mundo visual que o rodeia, de modo idêntico ao ser humano. Uma das características mais interessantes da Visão por Computador prende-se com a variedade de aplicações na vida real, tais como entretenimento, condução autónoma, saúde, segurança (reconhecimento facial), desporto, entre outros [1]. Uma aplicação muito específica deste tipo de tecnologia refere-se ao reconhecimento de gestos, nomeadamente no campo da acessibilidade [2]. Sendo o domínio desportivo e o reconhecimento de gestos duas aplicações específicas de *Visão Computacional* e *Aprendizagem Profunda*, o presente trabalho tem como objetivo a busca e construção de um modelo de classificação da sinalética gestual aplicada em competições de Karaté, na disciplina de Kumite, mais especificamente dos gestos do árbitro para atribuição de pontos, podendo a investigação realizada ser o ponto de partida para a produção de um sistema que torne o processo de arbitragem mais eficiente e objetivo.

Os gestos do árbitro para atribuição de pontos são: (1) *Yuko*: um ponto; (2) *Waza-ari*: dois pontos; e (3) *Ippon*: três pontos. Estes são cruciais no Karaté, sendo que a distinção entre os gestos para Aka (atleta à direita do árbitro) e Ao (atleta à esquerda do árbitro) é determinada pela direção em que são realizados, pelo que o modelo terá de classificar um total de 6 gestos. Apresenta-se, então, a investigação realizada na busca por um possível modelo de classificação da sinalética gestual do árbitro do Karaté, onde foram efetuadas três diferentes abordagens, todas tendo como base o recurso a modelos CNN (Convolutional Neural Networks – Redes Neuronais Convolucionais), ou baseados neste tipo de rede neuronal.

Identify applicable sponsor/s here. If no sponsors, delete this text box. (sponsors)

2024 19th Iberian Conference on Information Systems and Technologies (CISTI) – Iberian Proceedings of CISTI 2024
25th – 28th June 2024, University of Salamanca, Salamanca, Spain
ISBN: 978-989-36481-1-7

II. Estado-da-arte

A. Reconhecimento de gestos

É inegável que humanos e máquinas trabalham e continuarão a trabalhar em conjunto, sendo que já se verificam muitos exemplos desta realidade, como é o caso das habitações inteligentes, veículos autónomos, trabalho colaborativo de robots na indústria, entre muitos outros exemplos. Esta relação homem-máquina apresenta benefícios quando aplicados em outros campos, tais como acessibilidade, saúde, desporto, entre outros. Existem, ainda, outros campos em que a sinalética gestual substitui a língua falada, como é o caso das forças policiais e militares, na aviação e desporto. No campo tecnológico, o reconhecimento de gestos estreitou a ligação homem-máquina, facto que levou a inovar esta interação, no sentido de um sistema ser capaz de reconhecer e entender os gestos humanos, permitindo ao homem a execução de comandos que acabam por substituir periféricos, como ratos e teclados.

Nos anos 70, foi criado um primeiro protótipo de luva, conhecido como *Sayre Glove* [3], desenvolvida por Thamos DeFanti e Daniel Sandim, no Instituto de Tecnologia de Massachusetts (MIT), tendo como base uma ideia original de Richard Sayre., inicialmente concebida para a realização de controlos deslizantes, no entanto acabou por inspirar o desenvolvimento de outras luvas, incluindo as utilizadas para o reconhecimento de gestos. Exemplo disto é a *Grimes Digital Data Entry Glove*. Com o avanço tecnológico, ao longo dos anos, as luvas foram sendo substituídas pela adoção de câmaras como meio de implementar aplicações de Interação Humano-Computador, facto que tornou mais natural o reconhecimento de imagens [4]. As diversas abordagens para a recolha de dados [5] e reconhecimento de gestos incluem o recurso a luvas com sensores, marcadores coloridos para a identificação dos dedos e palmas das mãos, para além da abordagem baseada em visão, que envolve o recurso a câmaras e outros equipamentos baseados em visão.

No domínio do reconhecimento de gestos, trabalhos como o de Liu et al. [6] propõem o recurso a Redes Neuronais Convolucionais para o reconhecimento de gestos da polícia de trânsito chinesa. Outro trabalho [7] propõe a deteção de humanos em tempo real e o reconhecimento de gestos para resgate com recurso a veículos aéreos não tripulados, sendo que este último (o reconhecimento de gestos) foi implementado com recurso a um modelo CNN de 12 camadas compilado com recurso a *Keras* e *TensorFlow*.

B. Visão Computacional e Aprendizagem Profunda no domínio desportivo - Karate

Ao contrário do que acontece no reconhecimento da linguagem gestual, onde apenas ocorre análise das mãos, no que se refere ao reconhecimento de ação humana, esta foca-se na totalidade do corpo. No domínio desportivo, este tipo de tecnologia apresenta como aplicações [8] o reconhecimento das posições de jogadores, extração da trajetória da bola, determinação da posse de bola, análise e avaliação da prestação no decorrer dos jogos (*coaching*), identificação das ações dos jogadores, classificação dos gestos do árbitro, entre outros.

No que se refere ao Karaté, e tendo em conta que a qualidade das técnicas realizadas depende do rigor e exatidão com que são replicados, tendo sempre como base técnicas originais, as tecnologias em estudo têm, igualmente, aplicabilidade. Em [9] é, por exemplo, apresentado um exemplo de aplicação de *Visão Computacional* com recurso a técnicas de *Aprendizagem Profunda*, uma plataforma que virtualiza um dojo físico e que proporciona aos praticantes de Karaté a prática da arte em qualquer lugar e, no que se refere a técnicas de *Aprendizagem Profunda* para o reconhecimento de gestos das mãos. Neste artigo foram testados diferentes tipos de CNN, tendo-se atingido valores de 98% de *accuracy.* Relativamente ao Kumite, uma das disciplinas da arte caracterizada pela explosividade dos movimentos e pela interação entre os dois participantes, Echevarria e Santos [10] exploraram a aplicação de um algoritmo de *pose estimation* para extração das características dos movimentos realizados em *Ippon Kihon Kumite*.

C. Visão Computacional e Aprendizagem Profunda no domínio desportivo - Arbitragem

A presença de um árbitro é comum em desportos como o futebol, basquetebol, boxe e, também, no karaté. No caso do futebol, em [11] propõe-se a realização da monitorização dos jogadores e árbitros com recurso ao reconhecimento pela cor da camisola, uma abordagem de reconhecimento de múltiplos objetos, com recurso a Redes Neuronais Convolucionais, em conjunto com outras estratégias, para deteção e classificação de jogadores, árbitro, bola e fundo. Igualmente, em [12] recorreu-se ao mesmo tipo de rede neuronal. Afastado do conceito de monitorização dos jogadores e árbitros, em [13] é apresentada uma proposta de identificação dos gestos do árbitro em jogos de Basquetebol, onde, para a criação do *dataset*, foram recolhidas imagens de três tipos de gestos. Neste, a métrica de avaliação usada (*accuracy*) apresentou valores de 95,6%. Segundo os autores, este valor é consistente com o verificado em trabalhos realizados por outros (pouco superiores a), podendo esta situação dever-se ao *dataset* reduzido e às características das imagens usadas.

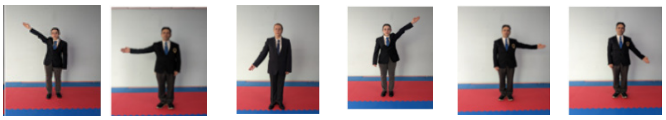
De um modo geral, e após se ter verificado a existência de vários domínios desportivos onde se pode recorrer às tecnologias em estudo, conclui-se que, no que se refere às artes marciais, os estudos existentes centram-se na classificação e reconhecimento de técnicas específicas de ataque e bloqueio, onde os objetivos se centram na elaboração de sistemas que suportem os praticantes das artes marciais, no entanto, não foram encontrados sistemas de reconhecimento da sinalética gestual de árbitros em contexto de Kumite.

III. Criação do dataset

A captação de fotos para a criação do *dataset* foi realizada em ambiente controlado (pavilhão desportivo/*dojo*), com diferentes condições de iluminação, tendo havido o cuidado de solicitar aos indivíduos que se vestissem conforme as regras da Federação Nacional de Karaté – Portugal [14], tendo sido captadas imagens de 4 indivíduos (2 do sexo feminino e 2 do sexo masculino), para um total de 6 gestos: (1) *Aka Ippon*; (2) Aka *Waza-ari*; (3) Aka *Yuko*; (4) *Ao Ippon*; (5) *Ao Waza*-ari; (6) Ao *Yuko* (fig.1). No total foram captadas 1038 fotos (173 por classe), sendo que a sua divisão respeitou as percentagens de 60, 20, 20 para os subconjuntos de treino, validação e teste, respetivamente (Tabela I). Foram, ainda, gravados 28 vídeos (7 por gesto) para testar com os modelos que apresentem potencial

2024 19th Iberian Conference on Information Systems and Technologies (CISTI) – Iberian Proceedings of CISTI 2024
25th – 28th June 2024, University of Salamanca, Salamanca, Spain
ISBN: 978-989-36481-1-7

para a realização de inferências em vídeo, tendo estes uma duração máxima de 3 segundos, com uma resolução de 900x900 e com 23,6 FPS.



Aka Ippon Aka *Waza-ari* Aka *Yuko* *Ao Ippon* *Ao Waza*-ari Ao *Yuko*

Figure 1. Exemplo de fotos capturadas para os 6 gestos do árbitro

Devido à dimensão reduzida do *dataset* houve a necessidade de se proceder a técnicas de *data augmentation*, como método de prevenção da ocorrência de *overfitting*. As transformações geométricas (Tabela II) aplicadas foram: *Rotation Range*, *Width Shift Range* e *Height Shift Range*, *Shear Range*, *Zoom Range* e *Horizontal Flip*. Como estratégia de preenchimento das lacunas deixadas pelas transformações anteriores (*Fill mode*), como por exemplo para casos em que ocorreram deslocamentos das imagens, optou-se pelo preenchimento com os pixéis mais próximos (*nearest*). Estas técnicas são aplicadas apenas ao subconjunto de treino, no entanto, todos os subconjuntos são sujeitos a técnicas de normalização e redimensionamento das imagens que os compõem. Após a aplicação destas técnicas de *data augmentation* o *dataset* aumentou para um total de 2946 fotos, e conforme se pode verificar na Tabela I, procurou-se criar um *dataset* equilibrado, onde a quantidade de imagens entre classes fosse igual, antes e após o *data augmentation*.

TABLE I. COMPOSIÇÃO DO DATASET

Classes	**Dataset**	**Dataset dividido**			**Dataset augmented**		
		Test	**Train**	**Valid**	**Test**	**Train**	**Valid**
Aka_ippon	173	33	106	34	33	424	34
Aka_waza_ari	173	33	106	34	33	424	34
Aka_yuko	173	33	106	34	33	424	34
Ao_ippon	173	33	106	34	33	424	34
Ao_waza_ari	173	33	106	34	33	424	34
Ao_yuko	173	33	106	34	33	424	34
Total	1038	1038			2946		

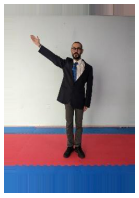
IV. METODOLOGIA

Seguidamente apresentam-se as três abordagens realizadas na investigação para a procura de um modelo de classificação da sinalética gestual do árbitro do Karaté.

A. Primeira abordagem: Classificação de imagens com recurso a Redes Neuronais Convolucionais

Com o objetivo de se proceder à classificação das imagens, foi construído, para esta abordagem, um modelo sequencial de uma Rede Neuronal Convolucional (*Conv2D*) composta por três camadas de convolução (conv2d_3, conv2d_4 e conv2d_5), seguidas por camadas de *pooling* (max_pooling2d_3, max_pooling2d_4 e max_pooling2d_5), que reduzem a dimensão de saída das camadas de convolução. Após as camadas de *pooling*, existe ainda uma camada *flatten* que transforma a saída das camadas anteriores num vetor unidimensional. Após esta, há duas camadas *dense* (dense_2 e dense_3) sendo que a última camada tem 6 neurónios, o que corresponde ao número de classes que o modelo deve classificar. Para o treino deste modelo a *framework* usada foi o *TensorFlow*, com recurso à API *Keras*.

TABLE II. TRANSFORMAÇÕES GEOMÉTRICAS APLICADAS AO DATASET

Foto original	**Transformações**	**Exemplos**
	rotation_range	
	width_shift_range	
	height_shift_range	
	shear_range	
	zoom_range	

B. Segunda abordagem: Modelo de deteção da pose humana com o modelo MoveNet

Chung *et al.* [15] apresentam uma análise comparativa de vários modelos de deteção da pose humana. Os autores analisaram os modelos *OpenPose*, *PoseNet*, *MoveNet* e *MediaPipe Pose*, tendo concluído que num *dataset* de imagens estáticas o modelo *MoveNet* revelou melhor performance na deteção de *keypoints* (pontos-chave) da postura humana e correta previsão da classe a que pertenciam, sendo que os resultados positivos verificados no modelo *MoveNet* foram igualmente satisfatórios quando aplicado a um *dataset* de vídeos. Perante estes resultados optou-se por avançar, nesta segunda abordagem, com o modelo *MoveNet*.

Para proceder à deteção da pose do árbitro e sua classificação, recorreu-se a uma implementação em TensorFlow do

2024 19th Iberian Conference on Information Systems and Technologies (CISTI) – Iberian Proceedings of CISTI 2024
25th – 28th June 2024, University of Salamanca, Salamanca, Spain
ISBN: 978-989-36481-1-7

modelo MoveNet para deteção de keypoints da postura humana em exercícios de Yoga [16], onde, para a sua classificação, é proposto um modelo Keras com duas camadas densas totalmente conectadas, com funções de ativação ReLu6 A camada de saída tem uma função de ativação Softmax, verificando-se a existência de um total de parâmetros treináveis.

O pré-processamento dos dados consiste na deteção das posturas nas fotos do dataset, para geração de ficheiros CSV com os keypoints de cada imagem, sendo este pré-processamento aplicado aos três subconjuntos. Os ficheiros gerados servem então de input para o modelo de classificação das posturas (fig.2).

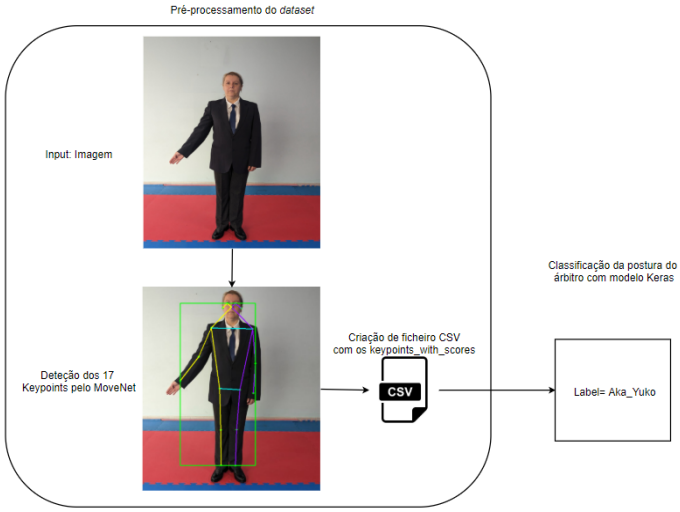


Figure 2. Pré-processamento de dados com *MoveNet*

C. *Terceira abordagem: Deteção e classificação de gestos com RoboFlow*

A plataforma *RoboFlow*, no seu plano gratuito, oferece várias opções de projeto, conforme o tipo de problema que se pretende resolver, e embora à primeira vista a opção de *Multi-Label Classification* pudesse parecer a solução ideal, a opção escolhida foi *Object Detection*, por permitir o uso em vídeo do modelo a treinar. Esta plataforma orienta o utilizador no sentido de iniciar o seu projeto com o *upload* do *dataset*, e neste ponto é importante referir que a mesma elimina fotos (aparentemente) duplicadas, pelo que o *dataset* usado para a construção do modelo nesta abordagem é de dimensão inferior ao usado nas abordagens anteriores (apenas 358 imagens). Por este motivo, as classes apresentam a distribuição da Tabela III, podendo afirmar-se que o *dataset* não é tão equilibrado como nas duas abordagens anteriores. No que se refere à sua divisão, a distribuição do mesmo foi de 250 imagens para o subconjunto de treino (70%), 71 imagens para o subconjunto de validação (20%) e 37 imagens para o subconjunto de teste (10%).

Após realização do Data Augmentation, o subconjunto de treino viu o seu número aumentar para um total de 750 imagens, passando o dataset para uma dimensão de 858 imagens, na sua versão aumentada. Das várias opções de Data Augmentation disponíveis na plataforma, as escolhidas foram: Exposure (between -6% and + 6%); Bounding Box: Exposure (between -12% and + 12%); e Bounding Box: Noise (up to 2% pixels).

O modelo escolhido para treino foi o YOLOv5 (You Only Look Once). Este é um tipo de modelo que apresenta uma arquitetura baseada em Redes Neuronais Convolucionais [17], amplamente utilizado para deteção de objetos em imagens, com grande capacidade de deteção em tempo real. No geral, a arquitetura deste modelo é composta por duas camadas totalmente conectadas e 24 camadas convolucionais.

TABLE III. Distribuição de imagens por classes (Roboflow)

Classes	**Aka Ippon**	**Aka Waza Ari**	**Aka Yuko**	**Ao Ippon**	**Ao Waza Ari**	**Ao Yuko**
# Imagens	59	55	64	59	59	60

V. Apresentação e Discussão de Resultados

A. *Primeira abordagem: Classificação de imagens com recurso a Redes Neuronais Convolucionais*

a) Treino do modelo Conv2D: O modelo inicial da primeira abordagem foi treinado por um total de 50 *epochs*, com a duração de 1592,02s (pouco mais de 26 minutos). Na última *epoch* o modelo apresentou loss de 0.93514 e accuracy de 1.000 para o conjunto de treino, significando que o modelo classificou corretamente 100% dos exemplos deste subconjunto. No que se refere ao subconjunto de validação, o modelo apresentou loss de 4.2243 e uma accuracy de 100%, concluindo-se que o modelo classificou corretamente todos os exemplos do subconjunto de validação.

b) Testagem e avaliação do modelo Conv2D: Após o treino do modelo, procedeu-se à sua avaliação no conjunto de teste, obtendo-se uma precision de 0.1970, recall de 0.1970 e accuracy de 19,70%, verificando-se que a aprendizagem do modelo não foi eficaz, situação reforçada por uma matriz de confusão que apresenta resultados aparentemente aleatórios (fig. 3).

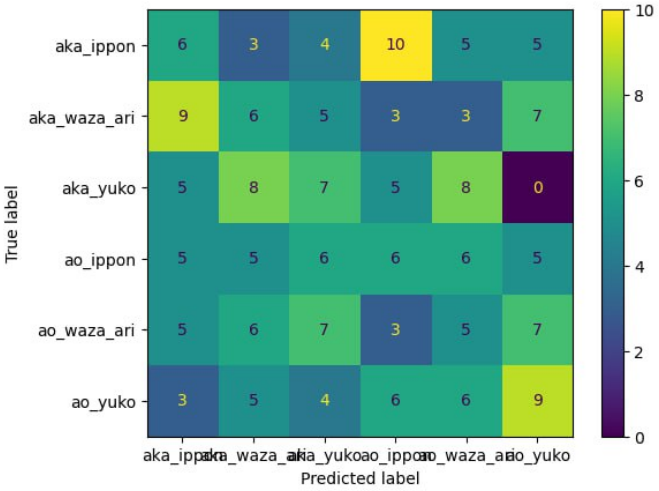


Figure 3. Matriz de confusão do modelo sequencial Conv2D

2024 19th Iberian Conference on Information Systems and Technologies (CISTI) – Iberian Proceedings of CISTI 2024
25th – 28th June 2024, University of Salamanca, Salamanca, Spain
ISBN: 978-989-36481-1-7

1) Transfer Learning e Fine-Tuning: Com o objetivo de melhorar os valores verificados após testagem do modelo sequencial Conv2D, recorreu-se a técnicas de *Transfer Learning* e *Fine-Tuning*. Para tal escolheram-se três modelos pré-treinados, disponibilizados pela API *Keras: VGG16, ResNet50 e Inception V3*. Depois a aplicação destas técnicas, verificaram-se os resultados obtidos na Tabela IV, após 50 *epochs* (10 com o modelo pré-treinado e 40 para o treino com *fine-tuning*), onde se observam resultados muito próximos aos obtidos com o Conv2D.

TABLE IV. RESULTADOS NOS MODELOS PRÉ-TREINADOS (TRANSFER LEARNING E FINE-TUNING)

		Modelos Pré-treinados		
		VGG16	**ResNet50**	**Inception V3**
Treino + Validação	Train_acc	1.000	0.9989	1.000
	Val_acc	1.000	1.000	0.9167
	Train_loss	0.0011	0.0144	0.0141
	Val_loss	1.0215	0.0300	0.1168
Teste	Accuracy	0.147	0.186	0.202
	Precision	0.1465	0.1869	0.2021
	Recall	0.1465	0.1869	0.2020

B. Segunda abordagem: Modelo de deteção da pose humana com MoveNet

1) Treino do modelo MoveNet: No dataset criado no âmbito deste trabalho, e aumentado como descrito anteriormente, este modelo apenas treinou durante 25 epochs (duração de 4.72s) por não se verificarem melhorias no valor da val_accuracy. O modelo em estudo atingiu, para o subconjunto de treino, uma accuracy de 0.9092 e loss de 0.3263. No subconjunto de validação o modelo atingiu uma accuracy de 1.000 e loss de 0.0123.

2) Testagem e avaliação do modelo MoveNet: Após se ter procedido à avaliação do modelo no subconjunto de teste verificou-se que este que atingiu uma precision e recall de 100%, indicando que o modelo fez previsões corretas para todas as amostras do conjunto de teste. Também a accuracy do modelo no conjunto de teste é de 100%, o que significa que o modelo classificou corretamente todas as amostras do conjunto de teste (fig.4). Mediante esta avaliação do modelo *MoveNet*, considerou-se estarem reunidas as condições necessárias para dar início à tentativa de se proceder à deteção da sinalética gestual do árbitro em vídeo.

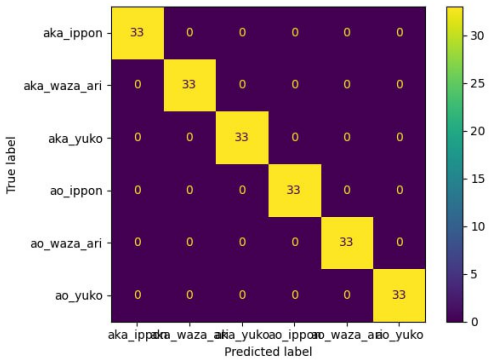
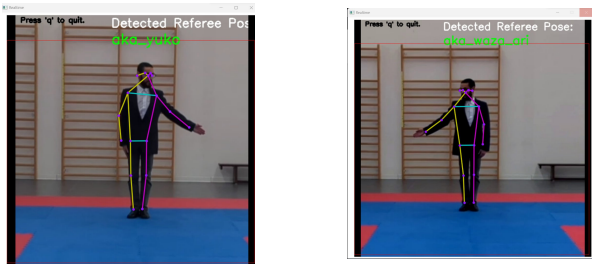


Figure 4. Matriz de confusão *MoveNet*

3) Classificação de posturas em vídeo com MoveNet: Na tentativa de testar o modelo para a realização de inferências em vídeo, frame a frame, o modelo de deteção de postura humana extrai os keypoints que, depois de normalizados, servem de input ao modelo proposto, implementado em TensorFlow e usado nesta segunda abordagem.

Esta inferência foi realizada apenas em 6 vídeos (1 por classe), onde se verificou que o modelo treinado, não só confundiu gestos como o lado para o qual estes ocorreram (fig. 5).



Classe real: Ao_yuko
Classe detetada: Aka_yuko

Classe real: Aka_yuko
Classe detetada: Aka_waza_ari

Figure 5. Exemplos do resultado da deteção em vídeo com recurso ao MoveNet

C. Terceira abordagem: Deteção e classificação de gestos com RoboFlow

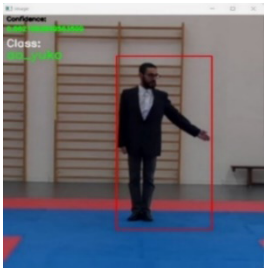
1) Treino, testagem e avaliação do modelo YOLOv5 no RoboFlow: O treino do modelo YOLOv5 no RoboFlow teve uma duração de 23 minutos, num total de 189 epochs (utilizador não controla o total de epochs), tendo atingido uma precision de 96.1%, onde as perdas para o treino foram reduzidas e a precisão elevada, logo após sensivelmente 50 epochs. Os resultados presentes na Tabela V, para os subconjuntos de validation e test, permitem afirmar que o modelo deverá ser capaz de generalizar bem para novos dados, podendo ser usado em segurança para fazer previsões em dados nunca vistos, pelo que este foi testado para a realização de inferências em vídeo.

TABLE V. PRECISÃO POR CLASSE DO MODELO YOLOV5 PARA OS SUBCONJUNTOS DE VALIDAÇÃO E TESTE.

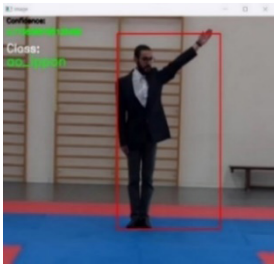
Classes	**Aka Ippon**	**Aka Waza Ari**	**Aka Yuko**	**Ao Ippon**	**Ao Waza Ari**	**Ao Yuko**
Validation set	1.00	0.94	1.00	0.98	1.00	1.00
Test set	1.00	1.00	1.00	1.00	1.00	1.00

Deteção classificação de gestos em vídeo com o RoboFlow: O RoboFlow não permite uso local do modelo criado, no entanto, é possível este ser usado como uma API personalizada [18] para realização de previsões em qualquer dispositivo com acesso à internet. Desta implementação observaram-se os resultados da fig. 6, onde se verifica que o modelo em estudo não confundiu nem gestos, nem os lados para o qual estes ocorrem.

2024 19th Iberian Conference on Information Systems and Technologies (CISTI) – Iberian Proceedings of CISTI 2024
25th – 28th June 2024, University of Salamanca, Salamanca, Spain
ISBN: 978-989-36481-1-7



Classe real: Ao_yuko
Classe detetada: Ao_yuko



Classe real: Aka_yuko
Classe detetada: Aka_yuko

Figure 6. Exemplos de deteção em vídeo com recurso ao modelo YOLOv5 (*RoboFlow)*

VI. Conclusões e trabalho futuro

Este trabalho centrou-se na busca pelo modelo ideal para classificação dos gestos que o árbitro de Karaté realiza, em contexto de competições, na disciplina de Kumite, para atribuição de pontos aos atletas. Para tal foram realizados três tipos de abordagens, onde os resultados obtidos em cada uma influenciaram as escolhas realizadas relativamente às abordagens seguintes. A primeira abordagem iniciou-se com a tentativa de classificar as imagens recorrendo a uma CNN (*Conv2D*). Nesta, o treino e validação apresentou resultados promissores, no entanto, quando se procedeu à testagem do modelo, verificou-se que a aprendizagem deste não foi eficaz, sendo esta situação reforçada pela matriz de confusão resultante, onde se observam valores aparentemente aleatórios. Com o objetivo de se obter melhores resultados, avançou-se para a aplicação de técnicas de *Transfer Learning* e *Fine-tuning*, no entanto os resultados obtidos na testagem dos três modelos pré-treinados (*VGG16*, *ResNet50* e *Inception V3*) foram próximos aos verificados no modelo base. Mediante estes resultados pôde concluir-se que depender exclusivamente da classificação de imagens com recurso a Redes Neuronais Convolucionais pode não ser a solução mais adequada para o problema em estudo. Uma vez que se estava perante um problema em que a postura humana é o cerne da questão, avançou-se para uma nova abordagem centrada em técnicas de *human pose estimation* através da deteção de *keypoints* que definem a postura do árbitro e consequente classificação das mesmas e com base na investigação realizada, optou-se por recorrer ao modelo *MoveNet*. Nesta segunda abordagem conseguiu-se obter resultados mais satisfatórios, uma vez que na testagem do modelo atingiram-se valores de 100% para *precision* e *accuracy*. Embora estes valores fossem promissores, verificou-se que na tentativa de realizar a deteção das posturas em vídeo o modelo não apresentou um resultado tão satisfatório, facto que poderá inviabilizar o recurso a esta solução para aplicação no mundo real. Por fim, e sempre com o objetivo de procurar o melhor modelo para a resolução do problema em estudo, avançou-se com uma terceira abordagem onde se recorreu à plataforma *RoboFlow* e ao treino de um modelo *YOLOv5,* com um *dataset* consideravelmente inferior. Nesta, foi atingida uma *precision* de 100% para todas as classes, tendo-se verificado que, de todos os modelos testados, este foi o que apresentou uma melhor performance no que se refere à deteção e classificação em vídeo, fazendo desta uma boa opção para a deteção e classificação da sinalética gestual do árbitro em competições de Karaté.

De futuro seria crucial investir na construção de um *dataset* de dimensão superior onde fossem incluídas, também, imagens de árbitros em contexto real de competição, com o objetivo de se proceder a um novo treino do modelo da última abordagem, para então este ser testado de modo intensivo para inferências em tempo real. Poder-se-á, após a obtenção de resultados mais sólidos no que se refere à inferência em vídeo, incluir mais sinalética gestual a ser classificada, como por exemplo para a identificação de faltas. Por fim, seria interessante o desenvolvimento uma solução que, após deteção da sinalética do árbitro, atualizasse as informações no placard de pontuações e faltas, e que pudesse ser integrado com algum do software já utilizado em contexto de competições.

Referências Bibliográficas

[1] What are the practical applications of computer vision technology? https://dac.digital/what-are-the-practical-applications-of-modern-computer-vision-technology/ (consultado a 04/04/2023)

[2] D. R. Antunes, C. Guimarães, L. S. García, L. E. S. Oliveira and S. Fernandes (2011). A framework to support development of Sign Language human-computer interaction: Building tools for effective information access and inclusion of the deaf, 2011 FIFTH INTERNATIONAL CONFERENCE ON RESEARCH CHALLENGES IN INFORMATION SCIENCE, Gosier, France, pp. 1-12, doi: 10.1109/RCIS.2011.6006832.

[3] D. J. Sturman and D. Zeltzer, (1994) *A survey of glove-based input*, IEEE Computer Graphics and Applications, vol. 14, no. 1, pp. 30-39. https://doi.org/10.1109/38.250916.

[4] B. Ma, W. Xu, and S. Wang. (2013) *A Robot Control System Based on Gesture Recognition Using Kinect*, TELKOMNIKA (Indonesian Journal of Electrical Engineering), vol. 11, no. 5, pp. 2605-2611. http://dx.doi.org/10.11591/telkomnika.v11i5.2493

[5] Al-Saedi, Ahmed & Al-Asadi, Abbas. (2019). *Survey of Hand Gesture Recognition Systems*. Journal of Physics: Conference Series. 1294. 042003. 10.1088/1742-6596/1294/4/042003.

[6] Liu, K., Zheng, Y., Yang, J., Bao, H., & Zeng, H. (2021). *Chinese Traffic Police Gesture Recognition Based on Graph Convolutional Network in Natural Scene*. *Applied Sciences*, *11*(24), 11951. MDPI AG. http://dx.doi.org/10.3390/app112411951.

[7] Liu, C., & Szirányi, T. (2021). *Real-Time Human Detection and Gesture Recognition for On-Board UAV Rescue*. *Sensors*, *21*(6), 2180. MDPI AG. http://dx.doi.org/10.3390/s21062180

[8] Naik, B. T., Hashmi, M. F., & Bokde, N. D. (2022). A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. *Applied Sciences*, *12*(9), 4429. MDPI AG. https://doi.org/10.48550/arXiv.2203.02281

[9] S. M. Jayasekara, S. S. Weerasinghe, D. Y. W. Abayawardana, A. R. Welagedara, S. E. R. Siriwardana and M. N. Koralalage. (2022) *Kaizen: Computer Vision Based Interactive Karate Training Platform*, TENCON 2022 - 2022 IEEE Region 10 Conference (TENCON), Hong Kong, Hong Kong, pp. 1-6. 10.1109/TENCON 55691.2022.9977691

[10] Echeverria, Jon, and Olga C. Santos. (2021). *Toward Modeling Psychomotor Performance in Karate Combats Using Computer Vision Pose Estimation*. *Sensors* 21, https://doi.org/10.3390/s21248378

[11] B. T. Naik, M. F. Hashmi, Z. W. Geem and N. D. Bokde, (2022) *DeepPlayer-Track: Player and Referee Tracking With Jersey Color Recognition in Soccer*. *IEEE Access*, vol. 10, pp. 32494-32509. *10.1109/ACCESS.2022.3161441*

[12] Guanghui Yang, Lijun Wang and Xiaofeng Xu et al. (2021*) Footballer Action Tracking and Intervention Using Deep Learning Algorithm*. Journal of Healthcare Engineering. 10.1155/2021/5518806

[13] Žemgulys, J., Raudonis, V., Maskeliūnas, R. *et al.* (2020) Recognition of basketball referee signals from real-time videos. *J Ambient Intell Human Comput* **11**, 979–991 https://doi.org/10.1007/s12652-019-01209-1

[14] Manual – Curso de Oficial de Mesa (Federação Nacional de Karaté – Portugal) https://www.fnkp.pt/wp-content/uploads/2020/06/Manual-Curso_Oficial_Mesa.pdf

2024 19th Iberian Conference on Information Systems and Technologies (CISTI) – Iberian Proceedings of CISTI 2024
25th – 28th June 2024, University of Salamanca, Salamanca, Spain
ISBN: 978-989-36481-1-7

[15] Chung, J.-L., Ong, L.-Y., & Leow, M.-C. (2022). *Comparative Analysis of Skeleton-Based Human Pose Estimation. Future Internet*, 14(12), 380. MDPI AG, doi: http://dx.doi.org/10.3390/fi14120380

[16] Classificação da postura humana com MoveNet e TensorFlow Lite https://www.tensorflow.org/lite/tutorials/pose_classification?hl=pt-br (consultado em 17/05/2023)

[17] Wu W, Liu H, Li L, Long Y, Wang X, Wang Z, et al. (2021) Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image. PLoS ONE 16(10): e0259283. https://doi.org/10.1371/journal.pone.0259283

[18] Hosted API (Remote Server) https://docs.roboflow.com/inference/hosted-api#displaying-the-response-image-with-format-image (acedido a 22/05/2023)

ISBN: 978-989-36481-1-7