CENTRO UNIVERSITÁRIO FEI
LAURA MARTINS VALENTE RAMOS BITTAR

AVALIAÇÃO DE MOVIMENTOS NA PATINAÇÃO ARTÍSTICA COM IDENTIFICAÇÃO E INTERPRETAÇÃO DE RESULTADO AUTOMÁTICAS

São Bernardo do Campo
2019

LAURA MARTINS VALENTE RAMOS BITTAR

AVALIAÇÃO DE MOVIMENTOS NA PATINAÇÃO ARTÍSTICA COM IDENTIFICAÇÃO E INTERPRETAÇÃO DE RESULTADO AUTOMÁTICAS

Dissertação de Mestrado apresentada ao Centro Universitário FEI, como parte dos requisitos necessários para obtenção do título de Mestre em Engenharia Elétrica. Orientada pelo Prof. Dr. Paulo Eduardo Santos.

São Bernardo do Campo

2019

Martins Valente Ramos Bittar, Laura.
Avaliação de movimentos na patinação artística com identificação e interpretação de resultado automáticas / Laura Martins Valente Ramos Bittar. São Bernardo do Campo, 2019.
97 f. : il.

Dissertação - Centro Universitário FEI.
Orientador: Prof. Dr. Paulo Eduardo Santos.

1. Patinação artística. 2. Aprendizagem profunda. 3. Algoritmos classificadores. I. Santos, Paulo Eduardo, orient. II. Título.

Elaborada pelo sistema de geração automática de ficha catalográfica da FEI com os dados fornecidos pelo(a) autor(a).

centro universitário FEI

APRESENTAÇÃO DE DISSERTAÇÃO
ATA DA BANCA EXAMINADORA

Programa de Pós-Graduação Stricto Sensu em Engenharia Elétrica

Mestrado

PGE-10

Aluno: Laura Martins Valente Ramos Bittar **Matrícula:** 116114-0

Título do Trabalho: Avaliação de movimentos na patinação artística com identificação e interpretação de resultados automáticas.

Área de Concentração: Inteligência Artificial Aplicada à Automação e Robótica

Orientador: Prof. Dr. Paulo Eduardo Santos

Data da realização da defesa: 17/12/2019

ORIGINAL ASSINADA

Avaliação da Banca Examinadora:

_____

_____

_____

_____

_____

São Bernardo do Campo, / / .

MEMBROS DA BANCA EXAMINADORA	
Prof. Dr. Paulo Eduardo Santos	Ass.: _____
Prof. Dr. Reinaldo Augusto da Costa Bianchi	Ass.: _____
Profª Drª Esther Luna Colombini	Ass.: _____

A Banca Julgadora acima-assinada atribuiu ao aluno o seguinte resultado:

APROVADO ☒ REPROVADO ☐

VERSÃO FINAL DA DISSERTAÇÃO

APROVO A VERSÃO FINAL DA DISSERTAÇÃO EM QUE FORAM INCLUÍDAS AS RECOMENDAÇÕES DA BANCA EXAMINADORA

_____

Aprovação do Coordenador do Programa de Pós-graduação

_____

Prof. Dr. Carlos Eduardo Thomaz

A Deus, por me permitir a vida e me guiar em cada momento da caminhada.

Aos meus pais e irmã, pelo incondicional amor, carinho e incentivo. Por estarem sempre de coração junto ao meu, não importa a distância.

AGRADECIMENTOS

A Deus, por mais uma etapa de aprendizado profissional e pessoal.

Aos meus familiares e amigos, pela paciência, apoio e força.

Ao meu orientador, professor Paulo Eduardo Santos, por compartilhar seu conhecimento, orientação e incentivo em todos os momentos.

Aos professores do Centro Universitário FEI, que desde o início dos meus estudos universitários contribuíram para a minha formação acadêmica e profissional.

Ao Centro Universitário FEI pelo apoio financeiro.

“Estude a ciência da arte. Estude a arte da ciência. Desenvolva seus sentidos – especialmente, aprenda como ver. Perceba que tudo se conecta a tudo mais.”

Leonardo da Vinci

RESUMO

A patinação artística é um esporte que apresenta forte combinação de elementos técnicos e artísticos. Seu sistema de avaliação envolve critérios objetivos, como o julgamento da execução de elementos técnicos, e subjetivos, como interpretação, composição e performance. O objetivo deste trabalho é apresentar uma proposta para identificação e avaliação automática de performances da patinação artística, inicialmente baseada nos critérios objetivos de julgamento do domínio, realizada por modelos artificiais. São utilizadas diversas técnicas de visão computacional, redes neurais artificiais e algoritmos classificadores para aprendizado supervisionado de máquina. Imagens são obtidas através de vídeos de competições oficiais de patinação artística, de forma a descrever em quadros a execução completa de elementos técnicos apresentados; a técnica de segmentação de imagens *graph-cut interativo* é aplicada nos quadros obtidos, de forma a separar o atleta do plano de fundo; imagens quadro-a-quadro são montadas para a criação de um conjunto de dados composto por 90 imagens com execução quadro-a-quadro de dois diferentes elementos técnicos (*Triple Axel* e *Quad Toe*); redes neurais profundas pré-treinadas (ResNet e Xception) são utilizadas para a extração de atributos de alto e baixo nível das imagens; por fim, modelos classificadores são treinados de forma supervisionada com os atributos previamente extraídos, possibilitando a identificação do elemento técnico apresentado e a avaliação obtida (pontuação acima ou abaixo do valor base do elemento). Através da combinação dessas técnicas e da avaliação de diferentes abordagens do problema, os modelos avaliados se mostraram capazes de identificar o elemento sendo realizado e avaliar a qualidade de execução dos elementos em relação aos seus valores base. Esta proposta visa contribuir para a ainda pouco explorada área de avaliação da qualidade de ações, atuando também como base e incentivo para desenvolvimento de futuros trabalhos que visem a aplicação de tecnologia para assistência de atletas, técnicos e jurados do esporte.

Palavras-chave: Patinação artística. Aprendizagem profunda. Algoritmos classificadores.

ABSTRACT

Figure skating is a sport with a great combination of technical and artistic elements. Its evaluation system involves objective criteria, such as the judgment of technical elements execution, and subjective criteria, such as interpretation, composition and performance. The aim of this study is to present a proposal for the automatic identification and evaluation of figure skating performances, initially based on objective criteria of the domain's judgment, performed by artificial models. Several techniques of computer vision, artificial neural networks and classifying algorithms for supervised machine learning are applied in this proposal. Images are obtained from videos of official figure skating competition, in order to describe in frames the complete execution of technical elements presented; the technique *interactive graph-cut* for image segmentation is applied to frames obtained, in order to separate the athlete from the background; frame-by-frame images are composed to create a dataset consisting of 90 images with frame-by-frame execution of two different technical elements (*Triple Axel* and *Quad Toe*); pre-trained deep neural networks (ResNet and Xception) are used to extract high and low level attributes from the images; finally, classifier models receive supervised learning through the previously extracted image attributes, enabling the identification of the performed technical element and the obtained evaluation (score above or below the element`s base value). Through these techniques' combination and the evaluation of different approaches to the problem, the evaluated models were able to identify the element being performed and to evaluate the elements execution quality in relation to their base values. This proposal aims to contribute for the still largely unexplored area of actions quality evaluation, also working as a basis and incentive for the development of future works that aims the application of technology to assist athletes, coaches and sports judges.

Keywords: Figure skating. Deep learning. Algorithms for classification.

LISTA DE ILUSTRAÇÕES

Figura 1 – Patins para patinação artística no gelo.....19
Figura 2 – Execução em “stop motion” (quadro a quadro) do salto Axel.....20
Figura 3 – Execução em “stop motion” (quadro a quadro) do salto Salchow.....20
Figura 4 – Execução em “stop motion” (quadro a quadro) do salto Toe loop.....21
Figura 5 – Execução em “stop motion” (quadro a quadro) do salto Loop.....21
Figura 6 – Execução em “stop motion” (quadro a quadro) do salto Flip.....22
Figura 7 – Execução em “stop motion” (quadro a quadro) do salto Lutz.....22
Figura 8 – Equipamentos necessários para eventos de alto nível na patinação artística.....24
Figura 9 – Ilustração de um graph-cut para a segmentação de imagem.....29
Figura 10 – Modelo de neurônio de McCulloch e Pitts.....31
Figura 11 – Exemplo de estrutura de uma rede neural convolucional.....35
Figura 12 – Um bloco de construção de aprendizagem residual.....37
Figura 13 – Uma versão canônica do módulo “Inception V3”.....38
Figura 14 – Uma versão “extrema” do módulo Inception, o módulo Xception.....39
Figura 15 – Arquitetura da rede Xception.....40
Figura 16 – Taxonomia hierárquica baseada em abordagens.....51
Figura 17 – Visão geral do modelo de agregação convolucional multiescalas com skip-LSTM.....58
Figura 18 – Diagrama esquemático da proposta apresentada nesta dissertação.....61
Figura 19 – Exemplos de imagens quadro a quadro montadas para o conjunto de dados.....64
Figura 20 – Arquitetura do modelo utilizado da rede ResNet-18 para Matlab, com 72 camadas.....68
Figura 21 – Arquitetura do modelo utilizado da rede Xception para Matlab, com 171 camadas.....70
Figura 22 – Abordagens utilizadas no treinamento de máquina para as tarefas de classificação.....72
Figura 23 – Matrizes de confusão na abordagem de classificação multiclasse, com atributos extraídos pela ResNet.....76
Figura 24 – Matrizes de confusão na abordagem de classificação multiclasse, com atributos extraídos para Xception.....76
Figura 25 – Matrizes de confusão na abordagem de classificação binária para identificação do elemento apresentado, com atributos extraídos pela ResNet.....79

Figura 26 – Matrizes de confusão na abordagem de classificação binária para identificação do elemento apresentado, com atributos extraídos pela Xception.....80
Figura 27 – Matrizes de confusão na abordagem de classificação binária para avaliação da qualidade de execução do Triple Axel, com atributos extraídos pela ResNet.....82
Figura 28 – Matrizes de confusão na abordagem de classificação binária para avaliação da qualidade de execução do Triple Axel, com atributos extraídos pela Xception.....83
Figura 29 – Matrizes de confusão na abordagem de classificação binária para avaliação da qualidade de execução do Quad Toe, com atributos extraídos pela ResNet....85
Figura 30 – Matrizes de confusão na abordagem de classificação binária para avaliação da qualidade de execução do Quad Toe, com atributos extraídos pela Xception.....86

LISTA DE TABELAS

Tabela 1 – Comparação de algoritmos de aprendizagem para classificação.....45
Tabela 2 – Resultados obtidos com a abordagem de classificação multiclasse.....75
Tabela 3 – Resultados obtidos com a abordagem de classificação binária para identificação do elemento apresentado.....78
Tabela 4 – Resultados obtidos com a abordagem de classificação binária para avaliação da qualidade de execução do Triple Axel.....81
Tabela 5 – Resultados obtidos com a abordagem de classificação binária para avaliação da qualidade de execução do Quad Toe.....84

SUMÁRIO

1 INTRODUÇÃO **13**

2 REVISÃO TEÓRICA **16**

2.1 PATINAÇÃO ARTÍSTICA 16

2.1.1 Modalidades de Competição Oficiais da Patinação Artística **17**

2.1.2 Saltos da Patinação Artística **18**

2.1.3 Sistema de Avaliação da Patinação Artística **23**

2.2 VISÃO COMPUTACIONAL PARA SEGMENTAÇÃO DE IMAGENS 26

2.2.1 Métodos de segmentação de imagem **26**

2.2.2 Técnica graph-cut para segmentação de imagens **27**

2.2.3 Categorias de algoritmos baseados em graph-cut **29**

2.3 REDES NEURAIS ARTIFICIAIS 31

2.3.1 Arquitetura e Características de Redes Neurais Artificiais **32**

2.3.2 Aprendizado de Redes Neurais Artificiais **33**

2.3.3 Redes de Aprendizagem Profunda **34**

2.3.4 Rede de Aprendizagem Profunda ResNet **36**

2.3.5 Rede de Aprendizagem Profunda Xception **37**

2.4 TÉCNICAS DE CLASSIFICAÇÃO PARA APRENDIZADO DE MÁQUINA 40

2.4.1 Atividades de classificação binária e multiclasse **41**

2.4.2 Algoritmos classificadores para aprendizado supervisionado de máquina **42**

2.4.3 Avaliação de algoritmos classificadores **45**

3 REVISÃO BIBLIOGRÁFICA **47**

3.1 RECONHECIMENTO DE AÇÕES HUMANAS BASEADO EM SISTEMAS DE VISÃO 47

3.1.1 Definição e aplicações **47**

3.1.2 Conjuntos de dados **48**

3.1.3 Representação de imagem **50**

3.1.4 Classificação de metodologias para reconhecimento de ações humanas 51

3.2 TRABALHOS RELACIONADOS 53

3.2.1 "Avaliando a Qualidade das Ações" - Pirsiavash, Vondrick e Torralba 53

3.2.2 "Aprendendo a pontuação em eventos olímpicos" - Parmar e Morris 55

3.2.3 "Aprendendo a pontuação de vídeos do esporte patinação artística" - Xu et al. 57

4 MÉTODO E IMPLEMENTAÇÃO 60

4.1 ABORDAGEM DO PROBLEMA E DETALHAMENTO DA PROPOSTA 60

4.2 ELABORAÇÃO DO CONJUNTO DE DADOS 62

4.3 EXTRAÇÃO DE ATRIBUTOS DAS IMAGENS ATRAVÉS DE APRENDIZAGEM PROFUNDA 65

4.4 CLASSIFICAÇÃO DAS IMAGENS ATRAVÉS DE APRENDIZADO DE MÁQUINA 71

5 ANÁLISE E DISCUSSÃO DOS RESULTADOS 74

5.1 CLASSIFICAÇÃO SEM HIERARQUIA ENTRE AS TAREFAS 74

5.2 CLASSIFICAÇÃO COM HIERARQUIA ENTRE AS TAREFAS – ELEMENTOS TÉCNICOS DIFERENTES 77

5.3 CLASSIFICAÇÃO COM HIERARQUIA ENTRE AS TAREFAS – QUALIDADE DE EXECUÇÃO DO TRIPLE AXEL 81

5.4 CLASSIFICAÇÃO COM HIERARQUIA ENTRE AS TAREFAS – QUALIDADE DE EXECUÇÃO DO QUAD TOE 84

5.5 DISCUSSÃO GERAL SOBRE OS RESULTADOS 86

6 CONCLUSÃO 90

REFERÊNCIAS 93

1 INTRODUÇÃO

A patinação artística é um esporte internacionalmente conhecido, que mescla elementos técnicos e critérios de avaliação objetivos, com elementos artísticos e critérios subjetivos, como interpretação, composição e performance. As características e estrutura do sistema de avaliação deste esporte permitem uma abordagem abrangente, detalhada e transparente das performances.

O objetivo deste trabalho é apresentar uma proposta para identificação e avaliação automática de performances da patinação artística, inicialmente baseada nos critérios objetivos de julgamento do domínio, realizada por modelos artificiais. Através da observação de diversas performances de elementos técnicos no domínio da patinação artística, o modelo deverá ser capaz de identificar o elemento sendo realizado e avaliar a execução em relação ao valor base do movimento, indicando posteriormente o elemento apresentado e se a performance foi acima ou abaixo do valor base.

Segundo Kotsiantis (31), a classificação supervisionada é uma das tarefas mais frequentemente realizadas pelos chamados sistemas inteligentes, visando a construção de um modelo conciso de distribuição dos rótulos das classes por meio de atributos preditores. O uso de algoritmos classificadores demanda a execução de algumas etapas prévias, como a coleta e montagem de um conjunto de dados, o pré-processamento dos dados visando melhorar a qualidade das informações relevantes ao problema e a identificação e seleção de atributos que representem de forma efetiva os dados (59; 60).

Para o desenvolvimento deste trabalho serão utilizados conceitos das áreas de visão computacional e inteligência artificial. Inicialmente, técnicas de segmentação de imagens serão aplicadas nas imagens observadas pelo agente artificial, de forma a possibilitar que ele capte e interprete os elementos da cena que sejam mais relevantes às atividades de classificação e avaliação que ele irá realizar.

A técnica de redes neurais pode ser utilizada como uma ferramenta para modelar relações complexas entre entradas e saídas de um processo ou para encontrar padrões em dados. Dessa forma, tendo uma quantidade de itens (performances de patinação artística) de diferentes classes (diferentes elementos técnicos e notas), a rede neural pode ser treinada e aprender a identificar padrões nesses dados, extraindo atributos que descrevam de forma relevante as características das imagens.

A realização de tarefas de classificação pode ocorrer através de diversos algoritmos classificadores existentes, baseados em diferentes técnicas. A escolha do algoritmo de

aprendizado a ser utilizado é uma etapa crítica do processo e depende sempre do problema a ser abordado. Assim, não existe uma forma fácil de determinar qual algoritmo é mais adequado para cada problema, sendo o teste de diversos algoritmos classificadores diferentes a maneira mais recomendada para encontrar a melhor solução.

Importantes e desafiadoras questões envolvem a proposta deste trabalho, tanto em sua etapa atual de avaliação dos critérios objetivos da performance, quanto em etapas futuras para avaliação dos critérios subjetivos. Os recentes avanços na área de visão computacional permitiram o desenvolvimento de diversos métodos para reconhecimento de ações em imagens e vídeos, porém a questão de avaliar o quão bem as ações foram executadas permanece largamente inexplorada (42). Dentre os diversos fatores que contribuem para a complexidade desta avaliação, podemos citar a dificuldade de encontrar e desenvolver bancos de dados que sejam específicos o suficiente para avaliar com qualidade as ações, contendo volume de dados adequado para um processo de treinamento efetivo.

Parmar e Morris (41), Pirsiavash, Vondrick e Torralba (42) e Xu et al. (54) desenvolveram importantes trabalhos na área de avaliação da qualidade de ações, apresentando propostas que abordam o tema por meio de estruturas genéricas baseadas no aprendizado de um modelo de regressão a partir de características espaço-temporais. Os domínios da patinação artística e saltos ornamentais foram utilizados para avaliação das propostas no mundo real, permitindo a identificação dos atuais desafios e indicando questões que precisam ser abordadas em pesquisas futuras para o desenvolvimento da área.

A proposta deste trabalho visa atuar como uma possível base para pesquisa e criação futura de tecnologia para assistência de atletas, técnicos e jurados do esporte, com aplicações no auxílio da avaliação e julgamento de performances, identificação de melhores sequências e sugestão aos atletas e técnicos para aprimoramento de elementos técnicos e artísticos, baseados na avaliação de performances do próprio atleta e de seus competidores.

A possibilidade de expandir e conectar diferentes áreas de conhecimento motivou fortemente a ideia desta pesquisa, que pretende contribuir em duas áreas ainda pouco exploradas da inteligência artificial: a avaliação da qualidade de uma ação e futuramente a abstração de etapas envolvidas no processo criativo humano, através das sugestões do agente artificial para melhorias das performances nos aspectos técnico e artístico.

Este trabalho apresenta, na seção 2, uma revisão teórica do domínio da patinação artística e das técnicas de inteligência artificial escolhidas para o desenvolvimento deste trabalho: visão computacional (técnicas de segmentação de imagens), redes neurais (arquitetura, características, aprendizado de redes e redes profundas) e técnicas de

classificação para aprendizado supervisionado (árvores de decisão, análise discriminante linear, naïve Bayes, k vizinhos mais próximos, máquinas de vetores de suporte, *ensembles* e regressão logística). Na seção 3, é realizada uma revisão bibliográfica que contempla conceitos das áreas de reconhecimento/qualidade de ações e apresenta trabalhos relacionados, que foram desenvolvidos com técnicas e domínios semelhantes aos escolhidos por esta proposta. A seção 4 apresenta de forma detalhada o desenvolvimento da pesquisa, demonstrando inicialmente um diagrama esquemático para auxiliar na compreensão da ideia e dos modelos/etapas envolvidos nesta proposta. Na seção 5 são apresentados e avaliados os resultados obtidos, com foco em identificar as questões abordadas que apresentaram bons resultados nesta proposta e as que necessitam de novas abordagens e desenvolvimentos futuros. A seção 6 apresenta a avaliação geral desta proposta e de trabalhos futuros.

2 REVISÃO TEÓRICA

A área de Inteligência Artificial utiliza um conjunto variado de técnicas para resolução de problemas, como Redes Neurais Artificiais (4; 47), Redes Bayesianas (5; 22) e Algoritmos Genéticos (19; 53), entre outras. Diversas dessas técnicas podem ser utilizadas para a criação de modelos de treinamento de máquina, tornando os algoritmos capazes de aprender informação diretamente dos dados, sem a necessidade de haver uma equação predeterminada como modelo (33).

As tarefas centrais da proposta dessa dissertação são a identificação e a avaliação dos elementos técnicos de uma performance de patinação artística. Ambas tarefas podem ser realizadas a partir de técnicas de aprendizado de máquina, pelo treinamento de algoritmos classificadores para a identificação de padrões em imagens de execuções dos elementos, tornando-os aptos a classificar itens não vistos anteriormente no treinamento.

Para o desenvolvimento deste trabalho serão utilizadas diversas técnicas: *visão computacional* (técnica de segmentação de imagens será aplicada nas imagens a serem observadas pelo algoritmo classificador), *redes neurais artificiais* (técnica utilizada para modelar relações complexas entre entradas e saídas de um processo ou para encontrar padrões em dados) e *algoritmos classificadores* (técnicas de classificação que possibilitam ao agente artificial atribuir classes à elementos anteriormente não vistos), com realização de *aprendizado supervisionado de máquina.*

Características do domínio da patinação artística, da visão computacional para segmentação de imagens, das redes neurais artificiais e de classificadores de aprendizado supervisionado de máquina, bem como suas aplicações específicas neste trabalho, serão descritas nas seções 2.1, 2.2, 2.3 e 2.4.

2.1 PATINAÇÃO ARTÍSTICA

A patinação artística é um grande exemplo da combinação entre esporte e arte. Desenvolvendo-se a partir de uma forma prática encontrada por nossos antepassados para se deslocarem sobre lagos e rios gelados, transformou-se em uma elegante mistura de movimentos artísticos e técnicas apuradas (39).

Estima-se que a patinação no gelo exista há 3 mil anos e que as primeiras combinações da patinação com a arte tenham surgido no século 18. Dois americanos foram responsáveis pelos maiores desenvolvimentos na história do esporte: em 1850, Edward Bushnell da

Filadélfia introduziu as lâminas de aço aos patins, o que permitiu a realização de manobras e giros complexos; na década de 1860, Jackson Haines, um mestre do ballet que vivia em Viena, incorporou música, coreografia e elementos de dança e ballet na patinação no gelo, dando início à patinação artística. As primeiras competições aconteceram na década de 1880 e, nos anos de 1908 e 1920, a modalidade fez parte dos jogos olímpicos de verão, passando a integrar os jogos olímpicos de inverno a partir de sua primeira edição, que ocorreu em 1924 em Chamonix – França. É o esporte mais antigo no programa dos jogos olímpicos de inverno e tornou-se um esporte muito popular, principalmente na América do Norte, Europa e Ásia, sendo disputado praticamente em todos os continentes (11; 39).

2.1.1 Modalidades de Competição Oficiais da Patinação Artística

A patinação artística pode ser disputada de forma *individual - masculino e feminino* (conhecido internacionalmente como "single program"), em *pares* ("pairs program") ou como *dança no gelo* ("ice dance"). Há também a modalidade *sincronizada* ("synchronized"), executada por equipes com 16 atletas, não sendo considerada, entretanto, como modalidade olímpica (25).

Conforme a União Internacional de Patinação - ISU (24; 25), dentre as principais características das modalidades olímpicas de patinação artística podemos citar:

- Individual: consiste em dois programas de patinação, o curto e o livre. O *programa curto* é composto pela execução de sete elementos obrigatórios, independente da ordem de execução, e de elementos de interligação. A duração total do programa não deve passar de dois minutos e quarenta segundos (2m e 40s). O *programa livre* é realizado com movimentos não obrigatórios, porém deve apresentar equilíbrio dos elementos técnicos (como saltos, piruetas, passos e elementos de ligação) e atenção na parte artística (coreografia, expressão e interpretação). A duração total máxima do programa livre deve ser de quatro minutos (4m).

- Pares: possui composição semelhante à do programa individual, porém as performances devem ser realizadas por duplas formadas por um homem e uma mulher, executando os elementos simultaneamente de forma simétrica (espelho) ou lado a lado. São adicionados nesta modalidade os elementos que só podem ser executados em dupla, como levantamentos, piruetas sincronizadas e saltos assistidos pelo parceiro. A duração total da performance não deve ultrapassar dois minutos e quarenta segundos (2m e 40s) no *programa curto* e quatro minutos (4m) no *programa livre*.

- Dança no gelo: é também composta por dois programas (*curto e livre*), porém nesta modalidade o casal é avaliado principalmente pelo elemento da dança. A avaliação é realizada com base na dificuldade dos elementos, na qualidade de execução e em critérios altamente subjetivos como entrosamento entre os parceiros, ritmo e harmonia. É a modalidade mais artística e, para garantir essa característica, há uma série de regras e exigências que o casal deve cumprir. São proibidos saltos, piruetas ou levantamentos sobre as cabeças, saltos e piruetas com mais de meia volta e os patinadores só podem permanecer separados durante alguns compassos musicais. Na modalidade dança no gelo a duração total da performance deve ser de dois minutos e cinquenta segundos (2m e 50s) para o *programa de dança curto* (ou dança rítmica) e quatro minutos (4m) para o *programa de dança livre*.

Na proposta deste trabalho serão avaliadas performances da *modalidade individual* – masculina e feminina. Na modalidade em pares, a variação dos posicionamentos dos atletas nas cenas é ainda maior do que na modalidade individual, já que há não somente a influência dos diversos posicionamentos de câmera e dos atletas em relação à pista, mas também dos dois atletas entre si. Já na modalidade de dança no gelo, a performance é realizada com foco no elemento da dança, sendo avaliada por meio de mais critérios artísticos, criativos e subjetivos do que nas modalidades individual e pares. Dessa forma, a modalidade individual foi escolhida por duas razões principais: diminuir a complexidade das cenas para identificação do agente artificial e obter cenas de elementos técnicos que sejam avaliados por meio de critérios objetivos, de forma a aumentar as chances da rede neural encontrar padrões nos dados.

2.1.2 Saltos da Patinação Artística

Os elementos técnicos da patinação artística na categoria individual podem ser classificados como saltos, rotações (ou piruetas), sequências de passos e sequências coreográficas. Existem diversos saltos na patinação artística, porém nem todos são considerados elementos oficiais das competições. Os seis saltos mais comuns podem ser divididos em duas categorias, sendo elas *"toe jumps"* (saltos com o dedo do pé) e *"edge jumps"* (saltos com os fios das lâminas dos patins). Dentre os toe jumps, estão os saltos *toe loop*, *flip* e *Lutz*. Dentre os edge jumps, estão os saltos *Salchow*, *loop* e *Axel*. Os saltos podem ser realizados com meia, uma, duas, três ou quatro rotações, possuindo valores de dificuldade diferentes de acordo com o número de rotações executadas (38; 52).

Todos os saltos possuem cinco componentes principais: preparação, posição de decolagem, salto no ar, rotação e aterrisagem. A observação desses componentes auxilia na identificação dos saltos, sendo a posição de decolagem o componente que mais difere entre os saltos. Um dos fatores importantes para a identificação dos elementos está relacionado aos patins e ao fio da lâmina que está sendo utilizado nos momentos da decolagem e aterrisagem. A figura 1 apresenta uma ilustração dos patins, na qual é possível identificar as lâminas internas e externas, além do serrilhado utilizado para o movimento *toe pick* (batida da parte da frente de um dos patins no gelo).

Figura 1 – Patins para patinação artística no gelo

Fonte: Autora "adaptado de" Edge Crunch (16)

Abaixo uma breve descrição dos seis saltos mais comuns, conforme a US Figure Skating, órgão nacional americano do esporte de patinação artística (52):

- *Axel*: nomeado em homenagem ao patinador norueguês Axel Paulsen, que foi o primeiro a realizar este salto. É considerado o salto mais difícil, sendo o único no qual a decolagem ocorre com o atleta patinando de frente (o que facilita na identificação do elemento) e sem auxílio de movimentos de preparação. O patinador segue para a frente utilizando o fio externo da lâmina do patim, em seguida salta com o auxílio da perna livre e necessita executar pelo menos uma rotação e meia no ar. A figura 2 apresenta uma ilustração quadro a quadro deste salto.

Figura 2 – Execução em “stop motion” (quadro a quadro) do salto Axel

Fonte: “adaptado de” Skate Sweden (48)

- *Salchow*: nomeado em homenagem ao patinador sueco Ulrich Salchow, que foi o primeiro a realizar este salto. Este salto possui decolagem com o atleta patinando de costas, no fio interno da lâmina do patim, e frequentemente é executado após movimentos de preparação. A perna livre é estendida para trás, balançando na sequência para a frente enquanto o patinador salta no ar, simultaneamente puxando os braços. O pouso é realizado no fio externo da lâmina do patim, com a perna oposta à que o atleta estava patinando na decolagem. A figura 3 apresenta uma ilustração quadro a quadro deste salto.

Figura 3 – Execução em “stop motion” (quadro a quadro) do salto Salchow

Fonte: “adaptado de” Skate Sweden (48)

- *Toe loop*: este salto possui decolagem com o atleta patinando de costas, no fio externo da lâmina do patim, com o auxílio do *toe pick* (batida da parte da frente de um dos patins no gelo). Frequentemente é executado após movimentos de preparação, sendo um dos saltos mais simples da patinação artística. O pouso é realizado também no fio externo da lâmina do patim, com a mesma perna que o atleta estava patinando na decolagem (perna oposta à utilizado para o *toe pick*). A figura 4 apresenta uma ilustração quadro a quadro deste salto.

Figura 4 – Execução em “stop motion” (quadro a quadro) do salto Toe loop

Fonte: “adaptado de” Skate Sweden (48)

- Loop: é o salto fundamental da patinação artística. Neste salto o atleta decola e pousa com a mesma perna e fio da lâmina do patim. A decolagem ocorre com o atleta patinando de costas, no fio externo da lâmina do patim, sendo frequentemente executado após movimentos de preparação. A figura 5 apresenta uma ilustração quadro a quadro deste salto.

Figura 5 – Execução em “stop motion” (quadro a quadro) do salto Loop

Fonte: “adaptado de” Skate Sweden (48)

- Flip**:** este salto se assemelha com o salto Salchow, porém com o auxílio do *toe pick* (batida da parte da frente de um dos patins no gelo). A decolagem ocorre com o atleta patinando de costas, no fio interno da lâmina do patim, sendo frequentemente executado após movimentos de preparação. O pouso é realizado no fio externo da lâmina do patim, com a perna oposta à que o atleta estava patinando na decolagem (perna utilizada para o *toe pick*). A figura 6 apresenta uma ilustração quadro a quadro deste salto.

Figura 6 – Execução em “stop motion” (quadro a quadro) do salto Flip

Fonte: “adaptado de” Skate Sweden (48)

- *Lutz*: nomeado em homenagem ao patinador austríaco Alois Lutz, é provavelmente o segundo salto mais famoso, após o Axel. Este salto possui decolagem com o atleta patinando de costas, no fio externo da lâmina do patim, com o auxílio do *toe pick* (batida da parte da frente de um dos patins no gelo). Frequentemente é executado após movimentos de preparação e de um longo tempo patinando de costas. É um salto contra rotacional (o atleta gira para um lado e salta para o outro), o que torna o elemento de difícil execução. O pouso é realizado no fio externo da lâmina do patim, com a perna oposta à que o atleta estava patinando na decolagem (perna utilizada para o *toe pick*). A figura 7 apresenta uma ilustração quadro a quadro deste salto.

Figura 7 – Execução em “stop motion” (quadro a quadro) do salto Lutz

Fonte: “adaptado de” Skate Sweden (48)

Na proposta apresentada nessa dissertação, execuções dos saltos Axel e Toe loop serão utilizadas para identificação e avaliação da qualidade do movimento por um agente artificial.

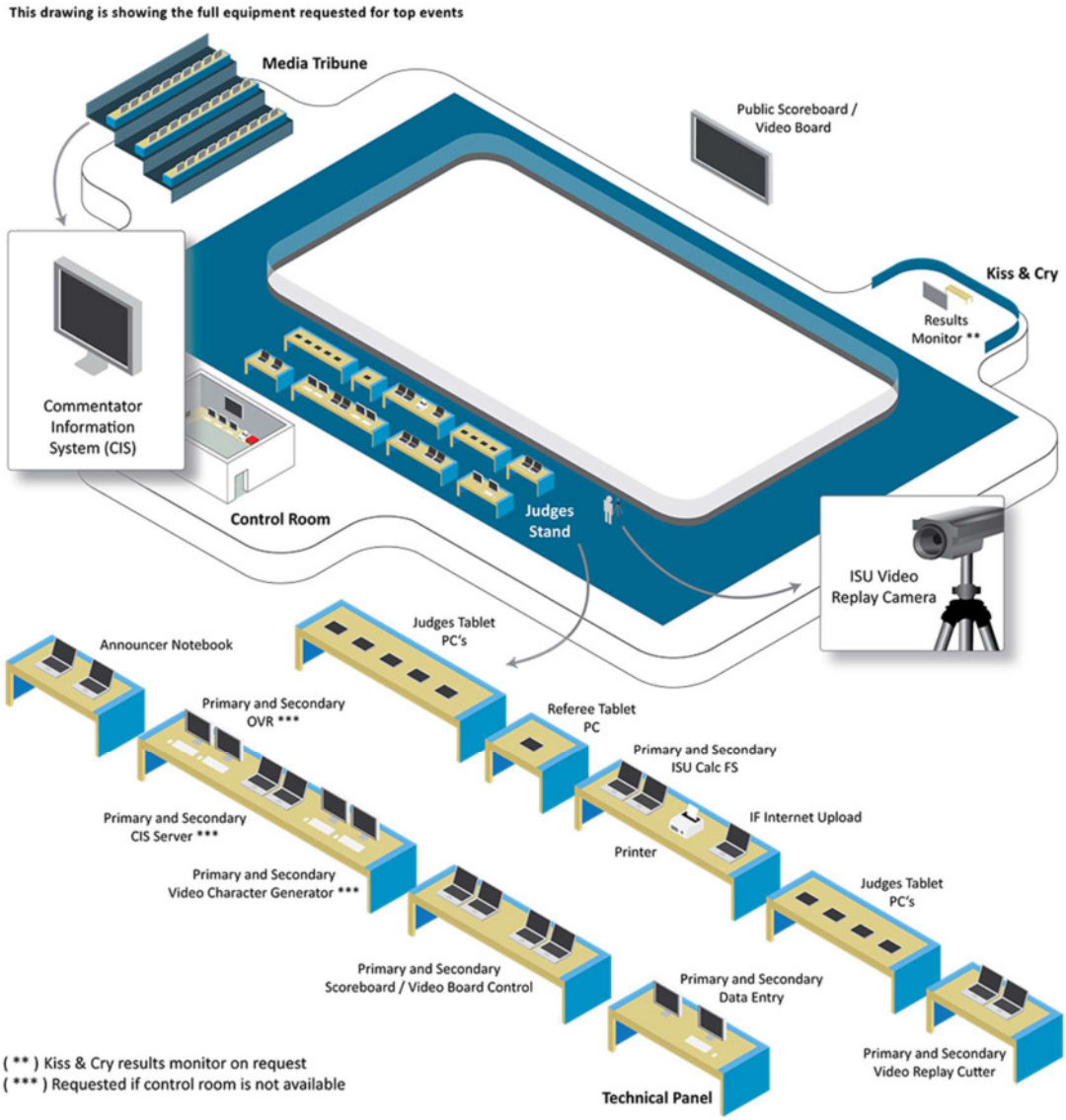
2.1.3 Sistema de Avaliação da Patinação Artística

A patinação artística teve seu sistema de avaliação totalmente remodelado após um escândalo nos jogos olímpicos de 2002, no qual avaliações corruptas e pontuações fixas vieram à tona, evidenciando a falta de critérios e objetividade do sistema anteriormente em vigor (sistema 6.0), o que facilitava a manipulação de resultados. O novo sistema desenvolvido, também chamado de "Código de Pontos", foi estabelecido para todas as competições oficiais e possui estrutura complexa, com diversos critérios que avaliam não apenas os elementos executados de forma individual, mas também a relação entre eles, como repetições, interligações, equilíbrio, momento de execução, entre outros. A nota final de uma performance é a combinação da pontuação da apresentação (maior foco nos componentes e critérios subjetivos), da pontuação técnica (foco nos elementos técnicos apresentados e critérios mais objetivos) e das deduções e/ou acréscimos.

O novo sistema implantado demanda uma estrutura composta por uma grande equipe técnica (controlador técnico, especialista técnico, assistente, operador de dados e operador de vídeo), além de sistemas computacionais específicos e um ambiente completamente conectado (25). Esta estrutura é também chamada de *painel técnico* e pode ser observada na figura 8. Dessa forma, é possível perceber que a patinação artística é um dos esportes que já utiliza diversas tecnologias computacionais para auxiliar na avaliação dos resultados e garantir maior confiabilidade ao processo.

Conforme a ISU (25), os jurados trabalham inteiramente focados em avaliar a qualidade dos elementos técnicos executados (pontuação técnica) e a qualidade da performance (pontuação da apresentação). A pontuação é baseada em critérios específicos de qualidade para cada elemento apresentado, avaliando o desempenho de cada atleta em relação aos critérios-base estabelecidos, e não em relação ao desempenho dos demais atletas.

Figura 8 – Equipamentos necessários para eventos de alto nível na patinação artística



Fonte: "adaptado de" Swiss timing (50)

As principais características das pontuações técnica e da apresentação são (24; 25):

- Pontuação técnica (TES – em inglês "total element score"): cada elemento executado possui um *valor base* e recebe uma nota chamada *grau de execução* (GOE – em inglês "grade of execution"), que varia em uma escala de -5 a +5, permitindo com que o valor base do elemento aumente ou diminua, de acordo com a qualidade de sua execução. A nota final para o *grau de execução* de cada elemento consiste na média das notas dadas pelos diferentes jurados, após a exclusão da maior e menor nota. Todos os elementos técnicos reconhecidos possuem um *valor base* e um *grau de dificuldade* previamente determinados, sendo o nível de pontos do valor base estabelecido de acordo com a dificuldade do elemento. Estes valores são periodicamente revisados por comitês técnicos e publicados anualmente em comunicado

oficial. Além do *valor base* e do *grau de execução*, também compõem a pontuação técnica valores atribuídos para *penalidades* e para *acréscimos*, conforme critérios e valores previamente estabelecidos no código de pontuação. A nota técnica final, ou TES, representa a soma das notas de todos os elementos executados.

- Pontuação da apresentação: a apresentação é avaliada de forma geral, através de cinco componentes do programa – *habilidade de patinação*, *transições*, *performance*, *composição* e *interpretação*. A componente *habilidade de patinação* avalia a precisão, segurança e clareza dos movimentos do atleta na pista de gelo, incluindo a capacidade de patinar com força nas acelerações, mudanças de velocidade e execução de elementos, enquanto demonstrando facilidade e pouco esforço. As *transições* avaliam a variedade, dificuldade e uso intencional de movimentos para interligação dos elementos. A *performance* inicia a avaliação de critérios mais abstratos, verificando o nível de envolvimento físico, emocional e mental do atleta, bem como sua projeção em relação ao público e à arena, os contrastes e a demonstração de sua personalidade por meio da apresentação. A *composição* avalia como o atleta usa a pista de patinação para a execução de seus elementos, o arranjo dos movimentos em relação à música e a existência de uma proposta ou ideia na estrutura da coreografia. Por fim, a *interpretação* avalia a tradução pessoal e criativa que o atleta realizou do conteúdo da música para os movimentos no gelo. As notas atribuídas variam em uma escala de 0,25 a 10, com incrementos de 0,25 pontos. Assim como para o grau de execução, a nota final para cada componente do programa consiste na média das notas dadas pelos jurados, após a exclusão da maior e menor nota. A nota de apresentação final, ou PCS, representa a soma das notas finais das componentes do programa, multiplicadas por um fator que depende do evento em questão.

A soma da nota técnica (TES), da nota de apresentação (PCS) e de possíveis deduções avaliadas compõe a nota total do atleta para aquela apresentação (TSS – em inglês "total segment score"). Para compor a nota final do atleta na modalidade, são somadas as notas obtidas nos programas curto e livre, sendo declarado como vencedor o atleta com maior pontuação na modalidade.

As características e estrutura deste sistema de avaliação permitem uma abordagem abrangente, detalhada e transparente das performances. A modelagem completa deste sistema por meio de um agente artificial seria uma tarefa de altíssima complexidade, que demandaria uma base de dados densa e de alta qualidade, de forma a permitir que o agente encontre padrões e compreenda as relações complexas entre os dados de entrada e saída do processo. Seria necessário, também, que o agente fosse capaz de avaliar os elementos tanto de forma individual, quanto como um conjunto (incluindo todos os elementos executados, a música e a

pista), para que fossem observados os critérios que relacionam os elementos entre si e avaliam os aspectos abstratos da apresentação. Dessa maneira, este trabalho adotou uma abordagem simplificada do problema, de forma a avaliar algumas propostas iniciais e propor novos trabalhos futuros.

As técnicas utilizadas para a abordagem proposta nessa dissertação serão descritas nas próximas seções.

2.2 VISÃO COMPUTACIONAL PARA SEGMENTAÇÃO DE IMAGENS

A segmentação de imagens é a realização do particionamento de uma imagem em diferentes regiões, que podem ter semelhança de cor, intensidade ou textura (20; 46). É uma etapa de pré-processamento que pode ser realizada por meio de técnicas de visão computacional e assume função importante em atividades de reconhecimento de objetos, rastreamento e análise de imagens (58).

Neste trabalho a segmentação de imagens foi utilizada para pré-processamento de imagens (quadros) extraídos de vídeos com performances de patinação artística. Na sequência será apresentada uma visão geral dos principais métodos de segmentação, bem como da técnica *Interactive-based graph cut*, utilizada na elaboração deste trabalho.

2.2.1 Métodos de segmentação de imagem

Segundo (58), as técnicas de segmentação de imagem são convencionalmente classificadas em cinco categorias, conforme apresentado abaixo.

- Segmentação baseada em limiares (em inglês "threshold") (20; 46): neste método, a imagem é normalmente dividida em duas partes, sendo elas o primeiro plano e o plano de fundo. A intensidade dos pixels da imagem é comparada com um limiar previamente definido e, quando os pixels possuírem intensidade maior/menor que este limiar, serão classificados como primeiro plano, caso contrário, serão classificados como plano de fundo. Diversos algoritmos se dedicam à tarefa de encontrar um limiar otimizado, sendo o algoritmo de Otsu (20) o método mais popular.

- Segmentação baseada em bordas (20; 58): este método assume que os valores dos pixels que compõem as áreas de divisão entre o primeiro plano e o plano de fundo são distintos. Dessa forma, utiliza métodos de derivadas de primeira ou segunda ordem, como gradiente ou Laplace, para detectar estas descontinuidades nos valores dos pixels e determinar as bordas

dos objetos. Normalmente, este método detecta também "falsas bordas", comumente necessitando de tarefas de pós-processamento para que seja obtido um resultado satisfatório.

- Segmentação baseada em regiões (20; 46): normalmente os algoritmos que se baseiam neste método abordam a segmentação pelo crescimento ou pela divisão de regiões. No caso do crescimento de regiões, é necessária a identificação de um conjunto de "sementes" a serem distribuídas pela imagem, de forma que os pixels vizinhos possam ser agrupados à essas sementes de acordo com critérios pré-estabelecidos como semelhança de intensidade, cor ou textura. No caso da divisão de regiões, a imagem é inicialmente dividida em uma série de pequenas regiões não sobrepostas, de forma que as regiões adjacentes que atenderem à critérios previamente estabelecidos serão posteriormente unidas novamente.

- Segmentação baseada no método Watershed (20; 45): este método faz uma analogia com o conceito topológico de bacias hidrográficas. Nele, as imagens são vistas como superfícies topológicas e os valores de intensidade como alturas. Os valores mínimos regionais são interpretados como bacias hidrográficas e os valores máximos entre bacias hidrográficas vizinhas são interpretados como linhas de cume. O método é normalmente aplicado no gradiente das imagens, de forma a tratar o objeto a ser segmentado como uma região de bacia hidrográfica e as suas fronteiras como linhas de cume. O método é sensível a ruídos e irregularidades no gradiente.

- Segmentação baseada em energia (20; 58): este método realiza a segmentação com base no estabelecimento de uma função de energia, que deve atingir um valor mínimo quando a imagem for segmentada conforme o resultado esperado. Diversas técnicas aplicam este método, como por exemplo a técnica *graph-cut* (8), utilizada para a etapa de segmentação das imagens na proposta dessa dissertação e descrita na seção seguinte.

2.2.2 Técnica graph-cut para segmentação de imagens

A segmentação graph-cut foi inicialmente proposta por Boykov e Jolly (8) em 2001. Nesta técnica, a função de energia é determinada com base em informações regionais e de fronteiras, podendo dessa maneira encontrar um resultado ótimo global. Desde sua proposta, foi base para o desenvolvimento de diversos algoritmos, sendo hoje uma técnica bastante aplicada em análise de imagens médicas, vídeos e segmentação de imagens naturais.

Conforme apresentado em (58), a base da técnica envolve um gráfico não direcionado indicado como $G =< V, E >$, no qual V é uma série de vértices (ou nós) e E é a aresta do gráfico que conecta dois vértices vizinhos. Os vértices V podem ser de dois tipos: *nós de*

vizinhança, que correspondem aos pixels, e *nós terminais*, que consistem em fonte (s) ou fundo (t). Também chamados de gráficos $s - t$, nestas imagens o nó s normalmente representa o objeto a ser segmentado (primeiro plano) e o nó t representa o plano de fundo. Há também dois tipos de arestas neste tipo de gráficos: as arestas *n-links*, que conectam os pixels vizinhos na imagem, e arestas *t-links*, que conectam os nós terminais com os nós de vizinhança. Cada aresta recebe um valor não negativo de peso (ou custo), denominado w_e. Um corte, denominado como C e expresso como $C \subset E$, é um subconjunto de arestas E, cujo custo $|C|$ pode ser determinado pela soma dos custos das arestas em C, conforme a equação 1.

$$|C| = \sum_{e \in C} w_e \tag{1}$$

Um corte que possua custo mínimo é denominado corte mínimo e pode ser obtido através de algoritmos que o identifiquem por meio do fluxo máximo para um dado gráfico $s - t$. Na segmentação de imagens, os rótulos dos pixels que compõem o objeto (nós s) são definidos como 1, enquanto os que compõem o plano de fundo (nós t) são definidos como 0, o que pode ser obtido pela minimização da função de energia, que ocorre através do corte mínimo do gráfico. Os cortes devem ser realizados nos limites entre o objeto e o plano de fundo, nos quais a energia deve ser mínima (58).

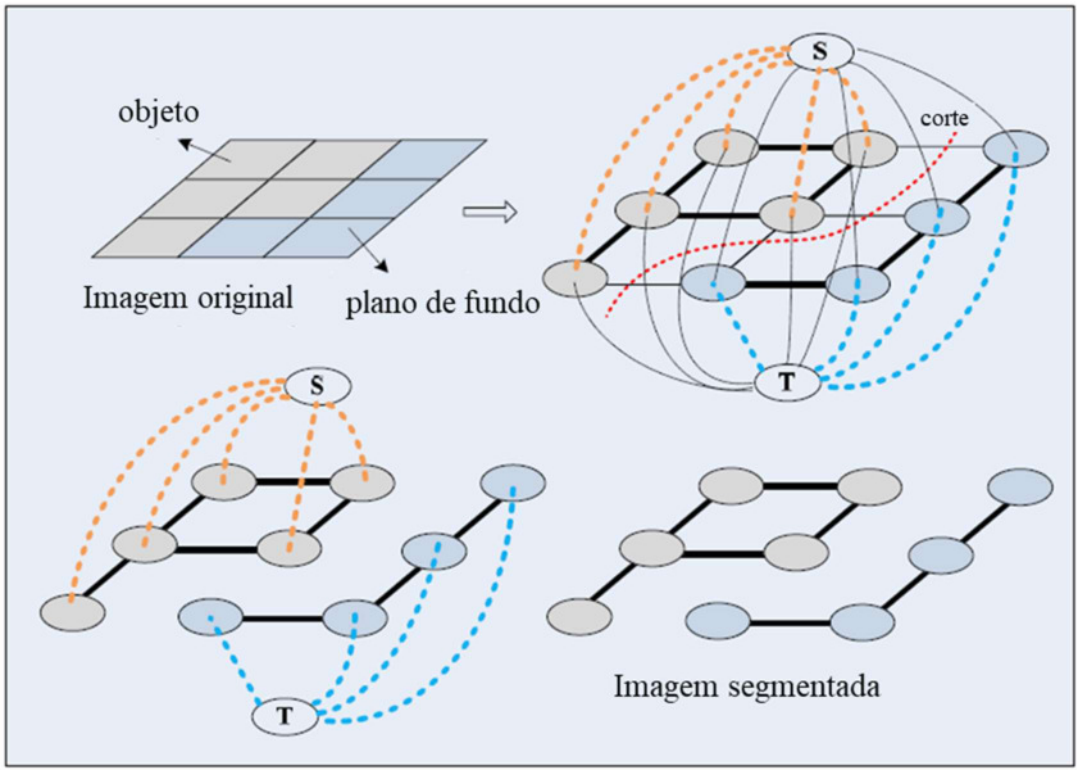
Seja $L = \{l_1, l_2, l_3, \dots, l_i, \dots, l_p\}$, onde p é o número de pixels na imagem e $l_i \in \{0,1\}$. Na realização do corte, o conjunto L é dividido em 2 partes, sendo que os pixels com rótulo 1 serão agrupados como objeto, e os pixels com rótulo 0 como plano de fundo. A função energia é definida na equação 2, na qual $R(L)$ é o termo que incorpora informação regional à segmentação, $B(L)$ é o termo que incorpora as restrições de fronteira e α é o fator de importância relativa entre os termos regional e fronteira. Detalhes da composição matemática dos termos da equação podem ser encontrados em (8).

$$E(L) = \alpha R(L) + B(L) \tag{2}$$

No método de Boykov e Jolly (8), quando um pixel apresentar tendências de pertencimento ao objeto, o peso (ou custo) da aresta entre esse pixel e o nó s será maior do que o da aresta entre o pixel e o nó t, o que significa que é mais provável ocorrer um corte na aresta com o menor custo. O mesmo ocorre com os pixels vizinhos, de forma que, quando o corte mínimo é alcançado no gráfico $s - t$, a localização do corte é próxima às fronteiras do objeto. A figura 9 ilustra o corte de um gráfico $s - t$ para a segmentação de uma imagem 3x3, na qual os pixels da imagem correspondem aos nós de vizinhança no gráfico (exceto os nós s e t), as linhas sólidas às arestas *n-links*, as linhas pontilhadas às arestas *t-links* e a espessura

das arestas representa a magnitude de seu peso. O corte apresentado corresponde à energia mínima, conforme equação 2.

Figura 9 – Ilustração de um *graph-cut* para a segmentação de imagem



Fonte: Autora "adaptado de" Yi e Moon (58)

Diversos algoritmos utilizam a técnica graph-cut como base, propondo também alguns ajustes, combinações ou etapas adicionais, visando aprimorar o resultado da segmentação. Esses algoritmos podem ser classificados em algumas categorias, conforme apresentado na próxima seção.

2.2.3 Categorias de algoritmos baseados em graph-cut

Yi e Moon (58) propõem a classificação dos algoritmos baseados em graph-cut em três diferentes categorias, sendo elas *graph-cut baseado em aceleração*, *graph-cut interativo* e *graph-cut baseado em forma prévia*, conforme descrição na sequência.

Os algoritmos *graph-cut baseados em aceleração* visam tornar o processo de segmentação mais rápido, seja por meio do uso de computação paralela ou através da redução gráfica. Nas abordagens de computação paralela, os algoritmos tendem a acelerar o processo de segmentação graph-cut através da implementação baseada no uso da GPU, apresentando

tempo computacional melhor do que na execução da computação sequencial. Nas abordagens através da redução gráfica, os algoritmos se baseiam na redução da quantidade de nós a serem avaliados durante o processo de segmentação e reconstrução da imagem. Na maioria dos casos, o objeto de interesse da segmentação ocupa apenas uma pequena parte da imagem como um todo, permitindo que seja selecionada uma área relativamente menor da imagem, reduzindo o tamanho do gráfico que deverá ser analisado e segmentado através do graph-cut. A seleção da região de interesse pode ser realizada pelo próprio usuário ou por meio do uso de algoritmos computacionais para seleção de regiões.

Os algoritmos *graph-cut interativos* visam melhorar a qualidade da segmentação nos casos em que a segmentação exclusivamente automática é complexa, ou quando uma acurácia alta é necessitada no processo. A segmentação interativa pode ser realizada tanto por meio da seleção de uma região de interesse, quanto pela seleção iterativa de pontos a serem utilizados como "sementes". A maioria dos algoritmos combina ambas as abordagens, de forma que após a seleção da região de interesse, pontos sementes devem ser selecionados, de forma a representar tanto o objeto quanto o plano de fundo, proporcionando ao gráfico uma melhor determinação dos pesos. A cada vez que um resultado da segmentação for apresentado e ainda não estiver de acordo com as expectativas, novos pontos sementes devem ser iterativamente adicionados, de forma a atualizar o processo e permitir que o algoritmo revise o resultado, até que seja atingida uma segmentação satisfatória do objeto.

Os algoritmos *graph-cut baseados em forma prévia* visam atuar nos casos em que há presença de ruídos na imagem, bordas pouco definidas ou objetos parcialmente oclusos. Nesses casos, a segmentação é complexa e o uso da abordagem interativa demanda muitas rodadas de iteração, afetando a eficiência da segmentação e nem sempre atingindo resultados satisfatórios. Dessa forma, os algoritmos incorporam uma informação prévia quanto à forma do objeto a ser segmentado na função de energia, conforme apresentado na equação 3, aprimorando com isso o processo e o resultado da segmentação.

$$E(L) = R(L) + B(L) + E_{forma} \tag{3}$$

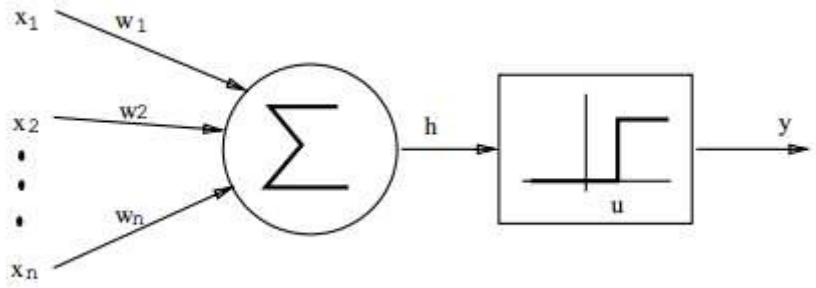
É importante ressaltar que as abordagens apresentadas nas três categorias acima não precisam ser necessariamente executadas de forma independente. Na maioria dos casos, os algoritmos utilizam uma combinação dessas abordagens, buscando melhorar o resultado de segmentação. As segmentações realizadas nas imagens deste trabalho utilizaram uma abordagem de segmentação graph-cut interativa, com seleção prévia de região de interesse e atualização dos resultados da segmentação por meio da seleção de pontos sementes.

2.3 REDES NEURAIS ARTIFICIAIS

Inspiradas nas redes neurais biológicas (grande quantidade de neurônios biológicos densamente conectados), as redes neurais artificiais (RNA) (4; 47) atuam como sistemas de processamento paralelo e distribuído, constituídos por uma grande quantidade de processadores simples (neurônios artificiais) e altamente conectados. Podem ser aplicadas para a resolução de problemas variados, como tomada de decisões, classificação de padrões, categorização, previsão, controle e otimização (27; 51).

O primeiro modelo matemático de neurônio artificial foi proposto por McCulloch e Pitts (35) na década de 1940, com o qual pretendiam demonstrar que programas da máquina de Turing poderiam ser implementados através de uma rede finita desses neurônios (27). Neste modelo, o neurônio artificial possui limiar binário (função degrau) e expressa a saída através de comparação da soma ponderada dos *n* sinais de entrada (x1.w1 + x2.w2 + ... = h) com o limiar estabelecido (u), apresentando valor de saída '1', caso a soma esteja acima deste limite, e ''0' no caso contrário, conforme ilustrado pela figura 10.

Figura 10 – Modelo de neurônio de McCulloch e Pitts



Fonte: "adaptado de" Jain, Mao e Mohiuddin (27)

O modelo de McCulloch e Pitts serviu como base para a expansão da área e foi generalizado de diversas maneiras, mostrando que múltiplas camadas de nós sem nenhuma inteligência de processamento (atuando como interruptores binários on-off) poderiam se aliar de forma autônoma através de pesos de conexões e da utilização de diferentes funções de ativação (função linear, sigmoide, gaussiana, tangentes hiperbólicas, degrau, etc.), eliminando a necessidade de se explicar manualmente, através de programação, regras ou cálculos dentro dos nós (27).

Por se tratar de técnica comumente utilizada para classificação de padrões e categorização (especialmente quando não se tem claramente compreendidas as relações complexas entre as entradas e saídas de um processo), as redes neurais artificiais foram escolhidas para serem utilizadas em parte do desenvolvimento deste trabalho, a fim de possibilitar a identificação e extração de atributos das imagens com os elementos apresentados nas performances de patinação artística. A estrutura de uma rede neural artificial e técnicas de aprendizado de redes serão abordadas nas seções seguintes.

2.3.1 Arquitetura e Características de Redes Neurais Artificiais

Uma Rede Neural Artificial (RNA) é constituída por um conjunto de neurônios artificiais de processamentos simples, interligados entre si e em camadas. Existem basicamente dois tipos de arquitetura de RNAs (56): *redes de alimentação direta* (feedforward) e *redes de retroalimentação* (feedback). As redes de alimentação direta são estáticas (dado um valor de entrada é produzido apenas um conjunto de valores de saída) e independentes do seu estado anterior (como não existem atualizações dos pesos das conexões, a resposta apresentada para dado valor de entrada não é influenciada pelo estado no qual a rede estava anteriormente). As redes de retroalimentação são dinâmicas (dado um valor de entrada é computado um novo valor de saída) e dependentes do estado anterior (por terem caminhos de retroalimentação, os pesos das conexões são atualizados e guardam, indiretamente, informações dos processamentos anteriores, levando a rede a entrar em novos estados).

Conforme Yadav, Yadav e Jain (56), dentre as principais características das RNAs podemos citar:

- Capacidade de processamento paralelo: o processamento paralelo biológico é altamente complexo, porém, por meio de métodos básicos e simples como matrizes e alguns cálculos, podemos realizar implementações simplificadas dessa capacidade em RNAs.

- Memória distribuída: por ser um sistema densamente conectado, uma memória local e centralizada não atende às exigências das RNAs, necessitando, assim, que dados sejam armazenados na matriz em forma de peso.

- Solução coletiva: as RNAs são sistemas interconectados, nos quais o resultado apresentado é o somatório de todas as saídas geradas após o processamento de várias entradas.

- Capacidade de tolerância a falhas: a falha de alguma parte das RNAs não afeta de forma impactante o sistema.

- Capacidade de aprendizagem: a atualização dos pesos das conexões permite que as RNAs possuam capacidade de aprender automaticamente com o decorrer do tempo, através das atividades que processam, se tornando eficientes na tarefa.

2.3.2 Aprendizado de Redes Neurais Artificiais

A capacidade de aprendizado é uma característica fundamental da inteligência humana e é também utilizada, ainda que de forma muito mais simplificada e generalista, nas RNAs. A partir de algoritmos de aprendizagem que se baseiam no ajuste dos pesos das redes, é possível que a rede consiga generalizar parâmetros por meio de uma coleção de exemplos representativos, ao invés de exigir do projetista a especificação de um grande conjunto de regras. Para determinar um algoritmo de aprendizagem adequado é necessário conhecimento do cenário do problema e ambiente no qual a rede opera (27).

Existem três formas principais de aprendizagem, sendo elas: *aprendizagem supervisionada* (a rede é provida com uma resposta correta para cada valor de entrada), *aprendizagem por reforço* (é uma variação da aprendizagem supervisionada, na qual a rede é provida apenas com o retorno quanto à exatidão da resposta apresentada e não com a resposta correta em si) e *aprendizagem sem supervisão* (a rede não requer nenhum tipo de resposta no conjunto de treinamento, explorando correlações nos dados e organizando padrões entre eles). (27; 56).

Conforme exemplificado por Jain, Mao e Mohiuddin (27), existem quatro tipos básicos de regras de aprendizado, sendo elas:

- Regras de correção de erros: o princípio da regra de aprendizado por correção de erro é utilizar os sinais de erro para modificação dos pesos das conexões, reduzindo o erro gradualmente. Um dos algoritmos mais conhecidos baseados no princípio de correção de erros é o algoritmo de retro propagação (backpropagation).

- Aprendizado de Boltzmann: é uma regra de aprendizagem estocástica derivada de princípios teóricos e termodinâmicos da informação. É um caso especial de aprendizagem por correção de erros, no qual o erro não é medido como a diferença entre a saída apresentada e a saída desejada, mas sim como a diferença entre correlações das saídas de dois neurônios.

- Regra de Hebb: é a regra de aprendizagem mais antiga, baseada na observação biológica de que se os dois neurônios conectados por dada sinapse forem ativados síncrona e repetidamente, então a força dessa sinapse é aumentada.

- Regra de aprendizagem competitiva: nesta regra, todas as unidades de saída são colocadas em competição umas com as outras para a ativação. Desta forma, apenas uma unidade de saída está ativa a cada momento e é, geralmente, a categorização dos dados de entrada.

Um algoritmo de treinamento deve ser adequadamente definido para que trabalhe de maneira equilibrada entre alcançar a estabilidade (não alterar o padrão dos dados de treinamento após dada quantidade de iterações) e a plasticidade (capacidade de se adaptar a novos dados).

Para realizar uma modelagem bem-sucedida em rede neural artificial, é fundamental um estudo criterioso de características como modelo e tamanho da rede, função de ativação, parâmetros de aprendizagem e tamanho do conjunto de treinamento. No entanto, ainda hoje, vários parâmetros precisam ser definidos por tentativa e erro. Uma das limitações (também potencial campo de pesquisa) da técnica de RNAs é que ainda não existe uma metodologia estruturada e padronizada para o desenvolvimento de uma RNA.

2.3.3 Redes de Aprendizagem Profunda

Redes profundas são redes neurais artificiais, cujo termo "profunda" refere-se ao número de camadas na rede, de forma que quanto mais camadas a rede tiver, maior sua profundidade. Redes neurais tradicionais possuem normalmente de 2 a 3 camadas, enquanto redes profundas podem ter centenas. A técnica de aprendizagem profunda é uma das técnicas de aprendizado de máquina, normalmente implementada utilizando uma arquitetura de rede neural profunda. A aprendizagem profunda é o estado da arte em aprendizado de máquina atualmente, principalmente devido a seu alto nível de acurácia (34).

O rápido desenvolvimento e expansão do uso de redes de aprendizagem profunda foi impulsionado principalmente por três fatores tecnológicos, sendo eles o fácil acesso a conjuntos vastos de dados rotulados (como o conjunto *ImageNet*, por exemplo), o aumento da capacidade computacional (principalmente através do uso de computação paralela com a GPU) e o desenvolvimento e disponibilidade de modelos pré-treinados construídos por especialistas da área (como os modelos AlexNet, ResNet, GoogleNet, entre outros).

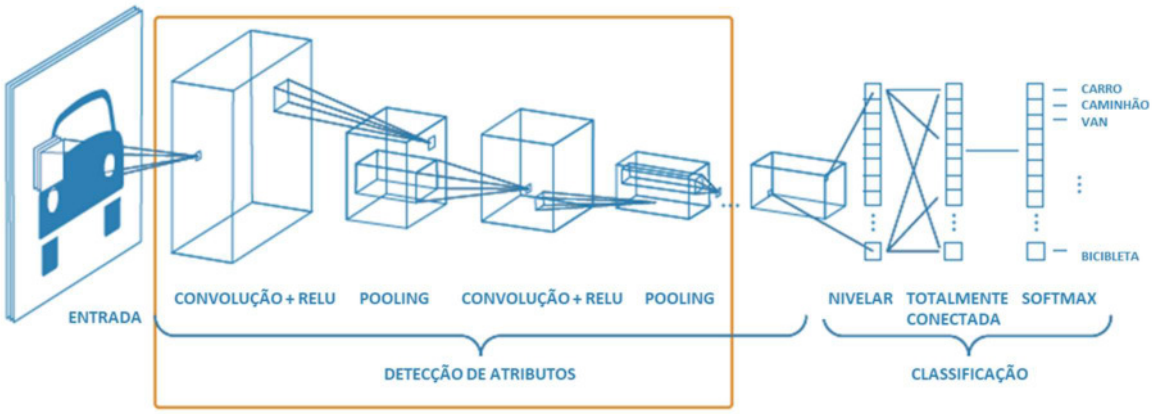
Na técnica de aprendizagem profunda, as redes aprendem a realizar tarefas de classificação diretamente dos dados de entrada, sejam eles imagens, vídeos ou sons. Dessa forma, não há influência sobre quais atributos dos dados estão sendo aprendidos, diferentemente das outras técnicas de aprendizado de máquina, nas quais a escolha dos atributos para descrição dos dados de entrada é uma etapa essencial, complexa e que afeta

diretamente a performance dos modelos. A aprendizagem profunda demanda milhares de dados para apresentar os melhores resultados, o que torna a técnica computacionalmente intensiva.

Rede neural convolucional (CNN, em inglês “convolutional neural network”) é um dos algoritmos mais populares para aprendizagem profunda com imagens e vídeos. Assim como outras redes neurais, a rede convolucional é composta de uma camada de entrada, uma camada de saída e muitas camadas ocultas no meio. Existem basicamente dois grupos de camadas: o grupo de detecção de atributos e o grupo de classificação.

As camadas do grupo de detecção de atributos realizam operações de convolução, pooling ou retificação linear dos dados. As camadas de convolução aplicam conjuntos de filtros nas imagens, cada um deles ressaltando determinados atributos das imagens. As camadas de pooling simplificam a saída através de reduções não lineares, de forma a reduzir o número de parâmetros que a rede necessita aprender. As camadas de retificação linear (ReLu, em inglês “rectified linear unit”) permitem um treinamento mais rápido e efetivo, através do mapeamento de valores negativos para zero e manutenção dos valores positivos. Estes três tipos de operações são realizados ao longo de diversas camadas, com cada uma delas aprendendo a detectar diferentes atributos. A figura 11 ilustra uma possível arquitetura de rede neural convolucional, destacando as camadas para detecção de atributos

Figura 11 – Exemplo de estrutura de uma rede neural convolucional



Fonte: Autora “adaptado de” Mathworks (34)

No grupo de classificação, a penúltima camada é totalmente conectada, cuja saída é um vetor de dimensão K, onde K é o número de classes que a rede poderá prever. Nesse vetor

estão contidas as probabilidades do dado de entrada pertencer a cada uma das classes. A última camada da rede fornece o resultado de saída da classificação.

2.3.4 Rede de Aprendizagem Profunda ResNet

Conforme apresentado na seção anterior, as redes profundas integram automaticamente atributos de baixo, médio e alto nível dos dados, além de estrutura classificadora, de forma que o processo é enriquecido pelo número de camadas (ou profundidade) da rede. No entanto, redes mais profundas frequentemente apresentam um problema de degradação, sendo que após começarem a convergir, a precisão fica saturada e na sequência degrada rapidamente. Essa degradação não é causada por um ajuste excessivo, mas sim pela profundidade da rede, de forma que a inclusão de camadas adicionais a um modelo de profundidade adequada leva a maiores erros no treinamento.

He et al. (21) propõem uma abordagem para o problema da degradação através da inclusão de uma estrutura de aprendizagem residual profunda, de forma que as camadas se ajustem apenas ao mapeamento residual e não diretamente ao mapeamento subjacente desejado. Para isso, são inseridas "conexões de atalho" de identidade, que não adicionam nenhum parâmetro extra e nem complexidade, mas transportam a informação de entrada diretamente através de uma ou mais camadas, permitindo que essas camadas empilhadas aprendam apenas a parte residual e garantindo que a informação de entrada não se perca completamente ao longo da rede profunda. Dessa forma, um anterior mapeamento através de uma função $H(x)$ é substituído por um mapeamento com $F(x) \coloneqq H(x) - x$ que, com a posterior adição da função identidade, reformula o mapeamento original em $F(x) + x$. A definição formal para o bloco de construção se dá conforme a equação 4, na qual x e F devem ter mesma dimensão.

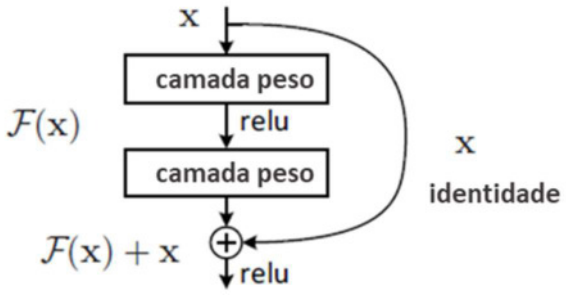
$$y = \mathcal{F}(x, \{W_i\}) + x \tag{4}$$

Caso a dimensão não seja a mesma, é possível realizar uma projeção linear W_s pela conexão de atalho, para corresponder as dimensões, conforme equação 5.

$$y = \mathcal{F}(x, \{W_i\}) + W_s\, x \tag{5}$$

A figura 12 exemplifica um bloco de construção da estrutura de aprendizado residual, em uma rede de alimentação direta (feedforward) com conexões de atalho.

Figura 12 – Um bloco de construção de aprendizagem residual



Fonte: Autora "adaptado de" He et. al, 2016 (21)

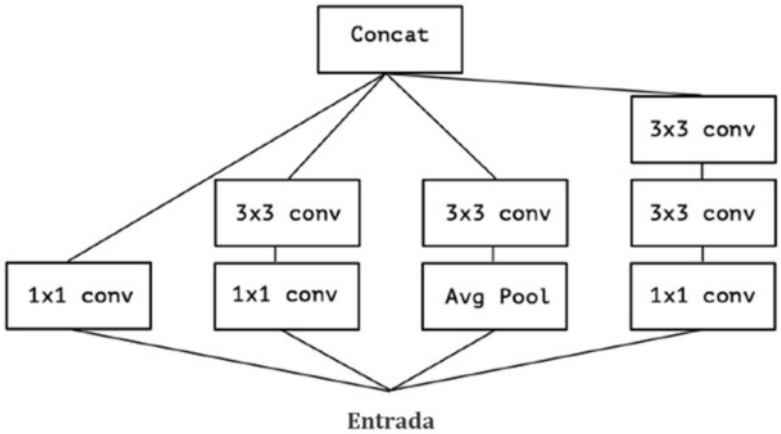
Essa estrutura de aprendizagem residual foi testada em diversos conjuntos de dados e com redes de diferentes profundidades, apresentando taxas de erro extremamente baixas e menores que a anterior arquitetura considerada estado-da-arte (VGGnet), ainda apresentando arquitetura de menor complexidade. Assim, a abordagem proposta para o problema da degradação em redes de aprendizagem profunda se mostrou eficiente e impulsionou diversos desenvolvimentos futuros.

2.3.5 Rede de Aprendizagem Profunda Xception

O desenvolvimento das redes de aprendizagem profunda vem apresentando grande expansão em curto tempo, especialmente inspirado pelos bons resultados que novas arquiteturas propostas estão alcançando. A rede Xception é uma arquitetura de rede neural convolucional profunda inspirada na rede Inception, na qual os "módulos Inception" foram substituídos por convoluções separáveis em profundidade, mantendo o mesmo número de parâmetros existentes na rede Inception V3, de forma que os ganhos de performance são devidos ao uso mais eficiente dos parâmetros e não ao aumento de capacidade.

Conforme Chollet (10), as redes Inception possuem um elemento fundamental, chamado de "módulo Inception", de forma que um modelo de rede Inception pode ser entendido como uma pilha desses módulos. Existem diversas versões para esse módulo, sendo a versão canônica encontrada na arquitetura Inception V3 apresentada na figura 13.

Figura 13 – Uma versão canônica do módulo “Inception V3”

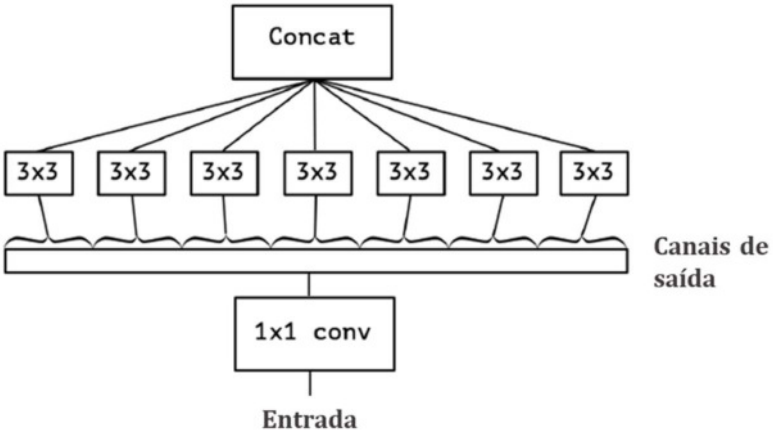


Fonte: Autora “adaptado de” Chollet, 2017 (10)

Os “módulos Inception” são conceitualmente parecidos com os módulos convolucionais, já que também realizam extrações de atributos convolucionais. No entanto, a arquitetura desses módulos buscou explorar o conceito do aumento de camadas em termos de largura, ao invés de apenas em profundidade. Os módulos convolucionais buscam aprender filtros em um espaço 3D, com duas dimensões espaciais (altura e largura) e uma dimensão de canal, de forma que um único núcleo realiza simultaneamente o mapeamento de correlações entre canais e de correlações espaciais. Os módulos Inception dividem este processo em uma série de operações que analisam independentemente as correlações entre canais e as correlações espaciais. Para isso, o módulo avalia inicialmente as correlações entre canais por meio de um conjunto de convoluções 1x1, mapeando os dados de entrada em 3 ou 4 espaços separados (menores que o espaço de entrada original), e mapeia na sequência todas as correlações nesses espaços 3D menores, por meio de convoluções 3x3 ou 5x5 (10).

Inspirado no módulo Inception, Chollet (10) propõe uma arquitetura baseada inteiramente em camadas de convolução separáveis em profundidade, com a hipótese de que o mapeamento de correlações entre canais e correlações espaciais nos mapas de atributos de redes neurais convolucionais pode ser totalmente separado. Com base nessa hipótese, uma versão “extrema” de um módulo Inception (a Xception), realiza primeiro uma convolução 1x1 para o mapeamento das correlações entre canais e, na sequência, mapeia de forma separada as correlações espaciais de cada um dos canais de saída. A figura 14 apresenta uma versão do módulo Xception, com uma convolução espacial por canal de saída da convolução 1x1.

Figura 14 – Uma versão “extrema” do módulo Inception, o módulo Xception



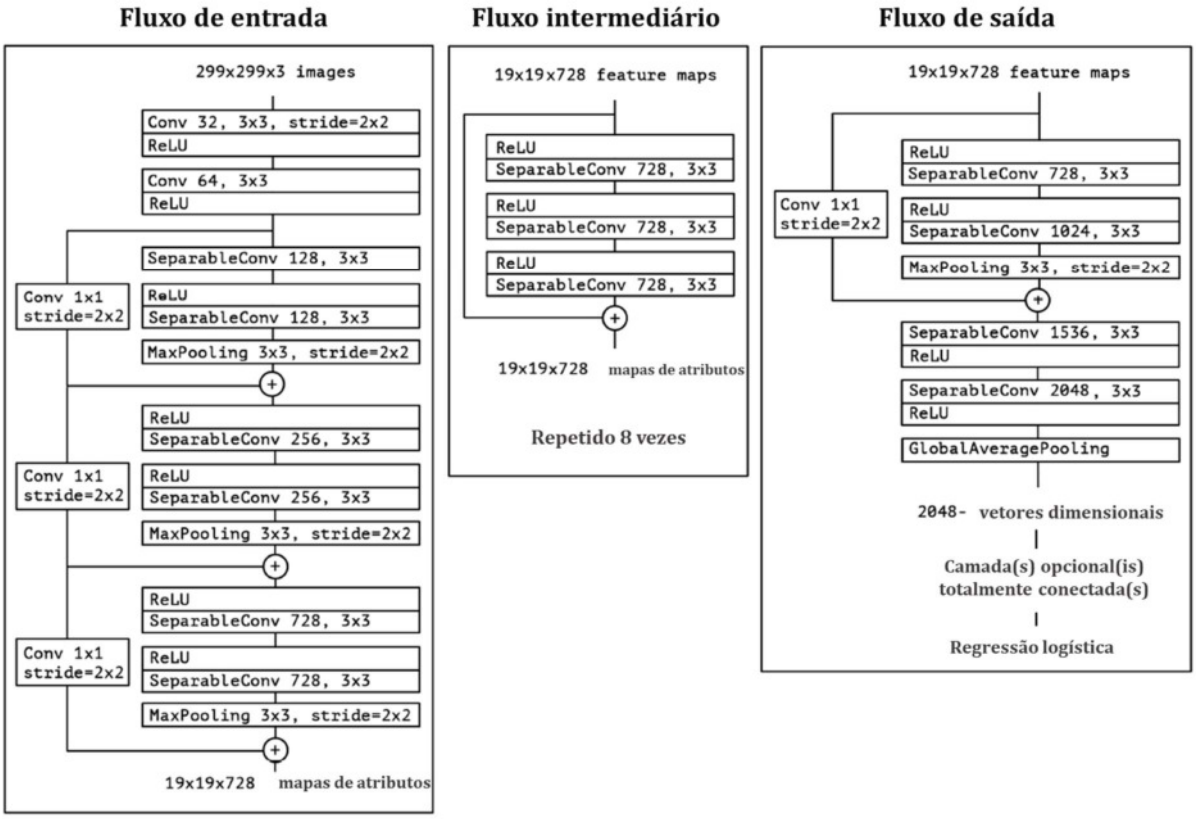
Fonte: Autora “adaptado de” Chollet, 2017 (10)

De uma forma simplificada, podemos interpretar essa arquitetura como o mapeamento de correlações em um espaço 2D inicialmente, seguido pelo mapeamento de correlações em um espaço 1D, o que permite um aprendizado mais simples do que através de um mapeamento 3D diretamente (55).

A arquitetura original da rede Xception possui 36 camadas convolucionais, compondo a base de extração de atributos da rede. Essas 36 camadas são estruturadas em 14 módulos, todos com conexões residuais lineares, com exceção do primeiro e último módulos. A figura 15 apresenta uma descrição completa da arquitetura e especificações da rede.

A arquitetura Xception foi avaliada em diversos conjuntos de dados, alcançando resultados ligeiramente melhores que a Inception V3 no conjunto ImageNet e significantemente melhores que a Inception V3 em um conjunto de dados maior (chamado JFT), composto de 350 milhões de imagens e 17.000 classes.

Figura 15 – Arquitetura da rede Xception



Fonte: Autora “adaptado de” Chollet, 2017 (10)
Legenda: Na arquitetura Xception, os dados passam primeiro pelo fluxo de entrada, depois pelo fluxo intermediário, que é repetido 8 vezes e, por último, pelo fluxo de saída.

Além da aprendizagem profunda existem diversas outras técnicas na área de aprendizado de máquina, com abordagens, características e aplicações diferentes. Algumas outras técnicas para classificação serão apresentadas na seção seguinte.

2.4 TÉCNICAS DE CLASSIFICAÇÃO PARA APRENDIZADO DE MÁQUINA

As técnicas de classificação (ou classificadores) para aprendizado de máquina possuem como objetivo principal reconhecer padrões nos dados de entrada que ofereçam ao algoritmo boas condições de realizar generalizações em dados não vistos, ao invés de apenas memorizar as informações dos dados utilizados no treinamento. (3)

Nesta seção são apresentadas formas de classificação, algoritmos classificadores comumente utilizados para aprendizado de máquina e métodos de validação da performance dos classificadores.

2.4.1 Atividades de classificação binária e multiclasse

As atividades de classificação podem ser divididas em duas categorias principais, de acordo com a natureza e abordagem do problema, sendo elas a *classificação binária* e a *classificação multiclasse*.

Na classificação binária, a tarefa envolve classificar elementos de um conjunto de dados em dois grupos, prevendo a qual dos dois grupos cada elemento de entrada pertence. Na classificação multiclasse, os elementos de um conjunto devem ser classificados em um de três ou mais grupos, de forma a apresentar resposta positiva para o grupo ao qual julga que a imagem pertença e resposta negativa para os demais grupos (2; 6).

O problema de classificação multiclasse é normalmente mais complexo, pois demanda dos algoritmos classificadores uma maior e melhor capacidade de generalização entre as várias opções. Alguns algoritmos classificadores naturalmente permitem, por sua forma construtiva, a realização direta de classificação multiclasse. No entanto, vários algoritmos usuais possuem natureza binária, ou classificação "um-contra-um", na qual uma das classes é tratada como a classe positiva e a outra como a classe negativa. Para permitir o uso de classificadores binários em problemas multiclasse, a maioria dos problemas podem ser abordados através das chamadas técnicas de transformação de problemas, que visam reduzir um problema de classificação multiclasse em múltiplos problemas de classificação binária, ou através das chamadas técnicas de adaptação de algoritmos, que visam estender os classificadores binários em classificadores multiclasse (2; 14).

As duas técnicas de transformação de problemas mais comuns para o caso da classificação multiclasse são as técnicas "um-contra-todos" e "um-contra-um". Na técnica *um-contra-todos*, uma das classes é tratada como a classe positiva, enquanto todas as outras são tratadas como a classe negativa, transformando o problema em um caso binário. No entanto, essa abordagem normalmente acaba apresentando um desbalanceamento entre a quantidade de dados de treinamento nas classes positiva e negativa, já que diversas classes foram agrupadas como negativas, podendo causar redução na performance de classificação. Na técnica *um-contra-um* são realizadas previsões de todas as combinações possíveis de pares de classes, transformando o problema em casos binários e mantendo um melhor balanceamento entre os conjuntos em cada uma das classes avaliadas (2; 6; 14).

2.4.2 Algoritmos classificadores para aprendizado supervisionado de máquina

Segundo Kotsiantis (31), a classificação supervisionada é uma das tarefas mais frequentemente realizadas pelos chamados sistemas inteligentes, visando a construção de um modelo conciso de distribuição dos rótulos das classes por meio de atributos preditores. Existem diversos algoritmos classificadores, baseados em diferentes técnicas, para a realização desta tarefa. A escolha do algoritmo de aprendizado a ser utilizado é uma etapa crítica do processo e depende sempre do problema a ser abordado. Assim, não existe uma forma fácil de determinar qual algoritmo é mais adequado para cada problema, sendo o teste de diversos algoritmos classificadores diferentes a maneira mais recomendada para encontrar a melhor solução.

O uso de algoritmos classificadores demanda a execução de algumas etapas prévias, como a coleta e montagem de um conjunto de dados, o pré-processamento dos dados visando melhorar a qualidade das informações relevantes ao problema e a identificação e seleção de atributos que representem de forma efetiva os dados (59; 60).

Um estudo abrangente sobre as diversas técnicas de classificação comumente utilizadas para aprendizado supervisionado foi realizado por Kotsiantis (31). Algumas dessas técnicas são utilizadas na proposta dessa dissertação, sendo apresentadas na sequência.

- ***Árvores de decisão:*** árvores de decisão são estruturas em forma de árvore que classificam instâncias com base nos valores dos atributos de representação do conjunto de dados. Cada um dos nós da árvore representa um atributo em uma instância a ser classificada, enquanto cada um dos ramos representa um valor que o atributo pode assumir. As instâncias são classificadas a partir de um nó raiz e categorizadas com base nos valores de seus atributos. A escolha do nó raiz é de fundamental importância, devendo ser escolhido para essa posição o atributo que melhor divide os dados do conjunto. A fim de evitar o sobre ajuste durante o treinamento é interessante podar a árvore antes que ela atinja um estado em que se ajusta perfeitamente a todos os dados do conjunto de treinamento, indicando uma possível memória dos dados. As árvores de decisão são uma técnica muito útil por terem forte característica de compreensibilidade, de forma que as pessoas podem facilmente entender por que uma instância foi classificada em uma classe específica. Normalmente apresentam resultados melhores ao lidarem com atributos discretos/categóricos.

- ***Análise discriminante linear***: é um algoritmo de aprendizado baseado em estatística, realizando a tarefa de classificação a partir da determinação da probabilidade de que uma dada instância pertença a cada uma das classes, selecionando a classe com maior

probabilidade para resultado da classificação. A análise discriminante linear é um método simples utilizado em estatística e aprendizado de máquina, para encontrar a combinação linear de atributos que melhor separe duas ou mais classes de elementos (18). É utilizada com valores contínuos, sendo equivalente à técnica de análise de correspondência discriminante, que é utilizada com valores categóricos (36).

- ***Naïve Bayes***: as redes bayesianas ingênuas são um dos mais conhecidos algoritmos baseado em estatística, sendo compostas de gráficos acíclicos direcionados com apenas um nó pai (nó não observado) e vários nós filhos (nós observados), assumindo forte independência entre os nós filhos no contexto do problema. O modelo de independência é baseado na estimativa apresentada na equação 6, através da qual a maior probabilidade indicará a qual classe o elemento mais provavelmente pertence. O classificador Naïve Bayes é normalmente menos preciso que outros algoritmos de aprendizado mais sofisticados, devido ao fato de que a suposição de independência entre os nós filhos está quase sempre errada. No entanto, uma das principais vantagens do classificador é o curto tempo computacional necessário para o treinamento (31).

$$R = \frac{P(i \mid X)}{P(j|x)} = \frac{P(i)\,P(X|i)}{P(j)P(X|j)} = \frac{P(i)\prod P(X_r\,|i)}{P(j)\prod P(X_r|j)} \tag{6}$$

- K vizinhos mais próximos (kNN, em inglês "k nearest neighborhood"): é um algoritmo de aprendizado baseado em instâncias, ou algoritmo de aprendizagem "preguiçosa", pois atrasa o processo de generalização até que a classificação seja realizada. Dessa forma, o algoritmo exige menos tempo de computação durante a fase de treinamento, porém mais tempo durante o processo de classificação. Segundo Cover e Hart (12), o algoritmo de k vizinhos mais próximos se baseia no princípio de que as instâncias constantes de um conjunto de dados geralmente existem em proximidade à outras instâncias que possuem propriedades semelhantes. Assim, o algoritmo kNN localiza as k instâncias mais próximas da instância de consulta e determina a sua classe de acordo com a classe mais frequente na vizinhança. A distância relativa é determinada através de uma métrica de distância, que idealmente deve minimizar a distância entre duas instâncias classificadas da mesma forma, e aumentar a distância entre instâncias de classes diferentes. Apesar de frequentemente apresentar boa precisão, dentre as principais desvantagens quanto ao uso de classificadores kNN podem-se citar a grande necessidade de capacidade de armazenamento, a dificuldade de escolher uma

métrica de distância adequada, a não existência de um princípio claro de definição para o parâmetro "k" e o longo tempo necessário para computação (31).

- Regressão logística: é um algoritmo de aprendizado baseado em estatística, para problemas de classificação binária. Realiza a tarefa de classificação através da análise estatística de um conjunto de dados, no qual exista uma ou mais variáveis independentes que determinem o resultado. A regressão logística visa encontrar o modelo mais adequado para explicar a relação entre a variável binária de interesse e uma ou mais variáveis independentes. Uma das dificuldades quanto ao uso da técnica de regressão logística é a dificuldade de interpretação do processo realizado pelo classificador (49).

- Máquinas de vetores de suporte (SVM, em inglês "support vector machines"): as máquinas de vetores de suporte baseiam-se na noção de margens e hiperplanos, sendo um método para problemas de classificação binária. O princípio do classificador é encontrar um hiperplano que separa duas classes de dados de forma a maximizar as margens, criando a maior distância possível entre o hiperplano de separação e as instâncias de ambos os lados. Para o caso de dados linearmente separáveis, uma vez encontrado o hiperplano de separação ideal, os pontos de dados que ficam na margem de ambos os lados ficam conhecidos como pontos do vetor de suporte, de forma que a solução é representada como uma combinação linear apenas desses pontos, ignorando os demais pontos de dados. Assim, a complexidade de um modelo SVM não está relacionada à quantidade de atributos encontrados nos dados de treinamento, mas sim ao número de vetores de suporte selecionados pelo algoritmo, que geralmente é pequeno. Essa característica de operação torna o classificador SVM adequado para lidar com tarefas de aprendizado nas quais o número de atributos é grande em relação ao número de instâncias de treinamento. Para o caso de dados não linearmente separáveis, uma solução possível é mapear os dados em um espaço dimensional superior (também chamado de espaço de recurso transformado) e, nesse espaço, definir um hiperplano de separação. Uma limitação do algoritmo é o longo tempo necessário para treinamento (9; 13).

- Conjuntos de classificadores (em inglês "ensembles"): a abordagem de combinação de classificadores é proposta como uma nova direção para melhoria do desempenho dos algoritmos classificadores individuais. Dessa forma, os classificadores *ensemble* são algoritmos de integração de resultados de classificação, visando utilizar os pontos fortes de um método para complementar os pontos fracos de outro, de forma a gerar resultados mais precisos. O estudo de métodos para construção de bons classificadores *ensemble* é uma área de pesquisa ativa em aprendizagem supervisionada. Os princípios frequentemente utilizados para a construção de classificadores *ensemble* são a utilização de diferentes subconjuntos de

dados de treinamento com um único método de aprendizagem, a utilização de diferentes parâmetros de treinamento com um único método de treinamento e a utilização de diferentes métodos de aprendizagem. Apesar de alcançar normalmente altos níveis de acurácia, entre as fraquezas dos classificadores *ensemble* podemos citar a necessidade de grande espaço de armazenamento, o aumento do tempo computacional e a diminuição da compreensibilidade quanto ao processo utilizado pelo classificador para encontrar uma resposta, já que com o envolvimento de vários classificadores na tomada de decisões se torna mais difícil a percepção do "processo de raciocínio" que efetivamente levou à decisão (31).

Kotsiantis (31) propõe uma tabela para comparação das características das técnicas de aprendizado, conforme apresentado na tabela 1.

Tabela 1 – Comparação de algoritmos de aprendizagem para classificação

	Árvores de decisão	Redes neurais	Naive Bayes	kNN	SVM
Acurácia em geral	**	***	*	**	****
Velocidade de aprendizado com relação ao número de atributos e ao número de instâncias	***	*	****	****	*
Velocidade de classificação	****	****	****	*	****
Tolerância a valores ausentes	***	*	****	*	**
Tolerância a atributos irrelevantes	***	*	**	**	****
Tolerância a atributos redundantes	**	**	*	**	***
Tolerância a atributos altamente interdependentes (por exemplo, problemas de paridade)	**	***	*	*	***
Lidar com atributos discretos / binários / contínuos	****	*** (não discretos)	*** (não contínuos)	*** (não direto discretos)	** (não discretos)
Tolerância ao ruído	**	**	***	*	*
Lidar com o risco de overfitting	**	*	***	***	**
Possibilidades/tentativas de aprendizado incremental	**	***	****	****	*
Explicação da capacidade / transparência do processo de raciocínio e classificação	****	*	****	**	****
Tratamento de parâmetros do modelo	***	*	****	***	***

Fonte: Autora "adaptado de" Kotsiantis (31), 2007
Legenda: **** representam o melhor desempenho e * representa o pior desempenho.

2.4.3 Avaliação de algoritmos classificadores

A avaliação da capacidade de generalização de um modelo classificador geralmente é realizada com base na acurácia de previsão. A técnica *validação cruzada* é amplamente empregada em problemas cujos modelos executam tarefas de previsão, buscando estimar o quão preciso é o modelo treinado para a avaliação de um novo conjunto de dados (30; 31).

O princípio central da técnica de validação cruzada é a divisão do conjunto de dados em subconjuntos mutuamente exclusivos, de forma que alguns destes subconjuntos serão posteriormente utilizados como dados de treinamento, para a estimativa dos parâmetros do modelo, enquanto os demais serão utilizados como dados de validação ou teste. A acurácia final estimada do modelo é dada pela equação 7.

$$Ac_f = \frac{1}{v} \sum_{i=1}^{v} \epsilon_{y_i,\hat{y}_i} = \frac{1}{v} \sum_{i=1}^{v} (y_i - \hat{y}_i) \tag{7}$$

onde v é o número de dados de validação e $\epsilon_{y_i,\hat{y}_i}$ é o resíduo dado pela diferença entre o valor real da saída i e o valor previsto.

Segundo Kohavi (30), existem diversas formas de realizar a divisão dos dados, sendo os métodos *holdout*, *k-fold* e *leave-one-out* os mais comumente utilizados.

O método *holdout* consiste em dividir o conjunto de dados em dois subconjuntos mutuamente exclusivos, um deles sendo utilizado para treinamento e outro para teste. A divisão do conjunto pode ser ou não em quantidades iguais, sendo uma proporção muito comum o uso de 2/3 dos dados para treinamento e 1/3 para teste. Esse método é indicado para quando existe disponível uma grande quantidade de dados.

O método *k-fold* consiste em dividir o conjunto de dados em *k* subconjuntos mutuamente exclusivos de tamanhos iguais, utilizando um deles para teste e os *k-1* restantes para o treinamento. Este processo é repetido *k* vezes, alternando o subconjunto de teste até que todos os subconjuntos criados já tenham sido utilizados uma vez para o teste. A acurácia do modelo é determinada através da média das acurácias de cada subconjunto testado.

O método *leave-one-out* é um tipo de *k-fold*, no qual *k* é o número total de dados *N*. Dessa forma, um subconjunto é criado para cada dado e, a cada uma das *N* repetições do processo, um deles é selecionado para teste e todos os demais para treinamento. A acurácia do modelo é determinada pela média das *N* acurácias calculadas, uma para cada dado. Esse método de validação é o que possui maior tempo computacional, sendo indicado para problemas com pouca quantidade de dados, ou quando é necessária uma estimativa mais rigorosa da acurácia do modelo (30).

As técnicas e conceitos apresentados neste capítulo foram utilizados no desenvolvimento desta pesquisa, conforme descrição e detalhamento a serem apresentados nos capítulos 4 (método e implementação) e 5 (análise e discussão dos resultados).

3 REVISÃO BIBLIOGRÁFICA

O domínio da patinação artística é extremamente abrangente, envolvendo conceitos técnicos de execução de movimentos, interpretação e expressão corporal, performance artística e criatividade. As apresentações e competições oficiais ocorrem em um ambiente com diversos recursos tecnológicos e computacionais, a fim de permitir uma melhor apreciação pelo público e criteriosa avaliação por parte das equipes técnicas.

A característica abrangente e multidisciplinar do domínio, ao mesmo tempo que possibilita a realização de pesquisas científicas de diversas áreas de conhecimento, apresenta também diversos desafios para sua modelagem, que deve encontrar formas de conectar todos os elementos envolvidos e abstrair informações subjetivas.

Os principais trabalhos técnicos encontrados neste domínio abordam a questão do reconhecimento de ações humanas baseado em sistemas de visão, combinando técnicas de visão computacional e inteligência artificial. Dessa forma, esta seção dedica-se a revisar os principais trabalhos na área de *reconhecimento de ações humanas baseado em sistemas de visão*, além de apresentar trabalhos relacionados que abordam o domínio da patinação artística e a avaliação de ações por meio do auxílio de técnicas de inteligência artificial.

3.1 RECONHECIMENTO DE AÇÕES HUMANAS BASEADO EM SISTEMAS DE VISÃO

3.1.1 Definição e aplicações

Segundo Poppe (43), o reconhecimento de ações humanas a partir de sistemas de visão refere-se ao processo de rotular sequências de imagens com rótulos que representam ações. De forma complementar, Jardim, Nunes e Dias (28) abordam o tema como um problema de classificação, no qual os eventos executados por humanos, apresentados em forma de vídeos (sequências de imagens), são automaticamente reconhecidos.

Este processo pode ser abordado por meio de diversas técnicas, podendo ser considerado, em sua forma mais simples, como uma combinação da *extração de atributos* que representem as imagens com subsequente *classificação* dessas representações. Assim, a abordagem comum consiste em primeiramente extrair atributos de imagens do vídeo que representa a ação, para atribuir, na sequência, um rótulo de classificação da ação identificada.

Os algoritmos de classificação são normalmente algoritmos de aprendizagem, treinados a partir de um conjunto de dados para treinamento (43).

Abordagens realizadas por Forsyth et al. (17) e Poppe (44) focam na recuperação de poses e movimentos humanos a partir de sequências de imagens, aproximando-se mais a problemas de regressão, enquanto o reconhecimento das ações humanas é tratado como um problema de classificação.

As diversas aplicações para o processo de reconhecimento automático de ações humanas a partir de sistemas de visão impulsionaram, na última década, um relevante crescimento e avanço das pesquisas, tornando este o tópico de pesquisa mais ativo nos campos de processamento de imagens e reconhecimento de padrões. Atualmente, as principais áreas de aplicação são sistemas de vigilância, interfaces humano-computador, recuperação de vídeos, análise de vídeos esportivos, assistentes de compras digitais, jogos e dispositivos para saúde e qualidade de vida (especialmente para idosos). Ainda assim, as tarefas de detecção de eventos em transmissões ao vivo ou em buscas de arquivos de vídeos continuam sendo majoritariamente realizadas por recursos humanos, o que enfatiza a complexidade da área e o ainda grande potencial de desenvolvimento e expansão (28; 43).

O reconhecimento de ações pode ser executado de diversas formas e em diferentes níveis de abstração. Uma importante estrutura foi proposta por Moeslund et al. (37), na qual as ações são tratadas de forma hierárquica e categorizadas como *primitiva de ação*, *ação* e *atividade*. Uma *primitiva de ação* é um movimento base, que pode ser descrito em termos de movimentos de membros. Uma *ação* é constituída por primitivas de ação, descrevendo um movimento de corpo inteiro, possivelmente cíclico. Uma *atividade*, por sua vez, é composta por um conjunto de ações subsequentes e descreve o movimento que está sendo realizado. Em uma abordagem semelhante, Bobick (7) utiliza uma estrutura para reconhecimento de *movimento*, reconhecimento de *atividade* e reconhecimento de *ação* que correspondem de forma aproximada às tarefas de visão de baixo, médio e alto nível.

3.1.2 Conjuntos de dados

Os conjuntos de dados são parte fundamental e extremamente importante dos processos de reconhecimento de ações humanas, sendo normalmente compostos por imagens ou vídeos. Atualmente é possível encontrar grande variedade de conjuntos de dados, porém a ausência de padrões na elaboração dos conjuntos e nas práticas utilizadas para avaliar as soluções propostas dificultam a comparação de diferentes trabalhos, permitindo apenas uma

comparação mais baseada nos conceitos do que nos resultados apresentados. Poppe (43) apresenta uma revisão sobre o tema e descreve as principais características e desafios dos conjuntos de dados com o foco de reconhecimento de ações humanas:

- *Variações nas execuções das ações*: diversas ações podem ser executadas de formas muito diferentes. O ato de andar, por exemplo, pode ser executado de formas diferentes por diferentes indivíduos (características biomecânicas de marcha) e até por um mesmo indivíduo (diferentes velocidades e passadas). Além disso, algumas formas de execução de determinada ação podem torná-la mais semelhante à alguma outra ação mapeada do que à ação originalmente executada. Uma boa proposta para reconhecimento de ações humanas deve ser capaz de, ao mesmo tempo, generalizar sobre as variações dentro de uma mesma classe e distinguir entre ações de classes diferentes.
- *Ambiente e configurações de gravação*: as características dos ambientes nos quais ocorrem as ações caracterizam grande fonte de variações nos dados. A localização de pessoas e identificação de movimentos se torna mais complexa em ambientes cheios, desorganizados, dinâmicos e com diferentes condições de luz. Detalhes do ambiente também podem encobrir partes das pessoas nas gravações, dificultando ainda mais as tarefas de reconhecimento.
- *Ângulos de visão e posições de câmeras*: uma mesma ação, observada por diferentes ângulos de visão, pode levar à observação de imagens muito diferentes. A maioria dos trabalhos apresentados se restringem à ângulos de visão fixos, o que diminui a variação no conjunto de dados, mas limita fortemente a aplicação das técnicas em situações reais. O uso de múltiplos pontos de visão e posicionamentos de câmeras torna as soluções mais aplicáveis e robustas, mas aumenta a complexidade do conjunto de dados e dos processos de treinamento.
- *Conjuntos de dados disponíveis publicamente*: grande parte dos trabalhos nesta área utilizam conjuntos de dados disponíveis publicamente, que foram especificamente criados para atividades de treinamento e avaliação, como os conjuntos *ImageNet* (15) e *COCO* (32). Ao mesmo tempo em que isto facilita a comparação entre as diferentes abordagens apresentadas, frequentemente estes conjuntos de dados não possuem as variações anteriormente citadas, o que torna os trabalhos mais distantes de uma aplicação em cenários reais. Alguns conjuntos de dados disponíveis já apresentam essas variações, porém são ainda limitados na quantidade de dados e sequências disponíveis, o que dificulta as atividades de treinamento.
- *Rotulagem de dados*: a rotulagem de dados se torna uma atividade desafiadora quando consideradas as diversas fontes de variações e normalmente demanda verificação manual. Com o aumento da quantidade e complexidade dos conjuntos de dados, a rotulagem manual

de dados se tornará algo cada vez mais difícil e custoso ao processo. Rotulagens automáticas já são possíveis em alguns domínios através de abordagens que utilizam as legendas e scripts dos vídeos, mas ainda demandam verificações manuais.

A área de reconhecimento de ações humanas teve grande evolução nas últimas décadas e, com isso, o nível dos algoritmos utilizados também evoluiu muito. Desta maneira, a criação de conjuntos de dados maiores e mais complexos demanda esforços de pesquisa da comunidade científica, a fim de que sejam criados conjuntos com configurações realistas. Atualmente já é possível observar uma maior tendência para a diversificação dos conjuntos de dados, o que traz benefícios não apenas à aplicabilidade das propostas, mas também ao desenvolvimento de conjuntos de dados para atividades mais específicas, ampliando os domínios de aplicação das técnicas (43).

Para o desenvolvimento da proposta deste trabalho, o conjunto de dados necessário é altamente específico (imagens quadro a quadro de elementos da patinação artística). Pirsiavash, Vondrick e Torralba (42) e Xu et al. (54) desenvolveram conjuntos que possuem vídeos com execuções de elementos de patinação artística – MITskate (42) e FIS-V (54), que poderiam ser utilizados neste trabalho como base para a criação das imagens quadro a quadro, reduzindo os esforços gastos na busca por vídeos e permitindo posteriormente a comparação de resultados entre as diferentes abordagens adotadas. Apesar das publicações citarem os conjuntos de dados como disponíveis publicamente, no período desta pesquisa não foi possível encontrá-los disponíveis de forma aberta na internet, gerando a necessidade do desenvolvimento de um novo conjunto de dados.

3.1.3 Representação de imagem

Conforme descrito nas seções anteriores, as imagens (ou sequências de imagens) que compõem um conjunto de dados são a base para o processo de reconhecimento de ações. Assim, é de fundamental importância a etapa de escolha e extração dos atributos de representação das sequências de imagens, de forma a garantir que as informações relevantes sejam transferidas para as próximas etapas do processo. De forma geral, a representação escolhida deve permitir a generalização sobre pequenas variações dentro da classe avaliada e, ao mesmo tempo, ser suficientemente robusta e precisa para conseguir identificar e avaliar ações de diferentes classes.

Segundo Poppe (43), a representação de imagens pode ser dividida em duas categorias, sendo elas:

- *Representação global*: descreve a observação visual como um todo. Nesta forma de representação, a região de interesse (ROI – em inglês "region of interest") é codificada como um todo e comumente obtida através de técnicas de rastreamento (em inglês "tracking") e subtração do plano de fundo, sendo derivadas de silhuetas, contornos/bordas e fluxo óptico. Essa forma de representação codifica uma grande quantidade de informação, mas também é sensível a ruídos nas imagens, oclusões parciais e variações nos ângulos de visão. Representações globais geralmente apresentam bons resultados e podem ser extraídas com baixo custo, porém possuem aplicação restrita aos cenários nos quais a região de interesse possa ser determinada com confiabilidade.
- *Representação local*: descreve a observação visual como um conjunto de descritores locais (ou partes) independentes. Nesta forma de representação, a localização precisa e a subtração do plano de fundo não são necessárias. É pouco sensível a variações nos ângulos de visão, nas aparências das pessoas e a oclusões parciais, sendo então uma alternativa confiável às situações nas quais a representação global não apresenta bons resultados.

As duas formas de representação de imagem acima apresentadas possuem caráter generalista quanto à diferentes domínios. Dependendo da pesquisa e situação, pode ser necessário e interessante utilizar formas de representação que sejam diretamente adaptadas de acordo com as características do domínio em questão.

3.1.4 Classificação de metodologias para reconhecimento de ações humanas

Aggarwal e Ryoo (1) sugerem uma interessante taxonomia baseada em abordagem, por meio da qual revisam diferentes trabalhos de pesquisa na área e as abordagens propostas para a realização do reconhecimento de ações. Uma ilustração com a visão geral da taxonomia pode ser encontrada na figura 16.

Figura 16 – Taxonomia hierárquica baseada em abordagens

Fonte: Autora "adaptado de" Aggarwal e Ryoo (1)

Os métodos de reconhecimento de ações são inicialmente classificados em duas grandes categorias, sendo elas as *abordagens de camada única* (em inglês "single-layered approaches") e as *abordagens hierárquicas* (em inglês "hierarchical approaches"). As abordagens de camada única representam e reconhecem ações humanas baseadas diretamente na sequência de imagens, sendo mais adequadas para o reconhecimento de gestos e ações com características sequenciais. As abordagens hierárquicas representam ações de alto nível, que são compostas e descritas em termos de outras ações mais simples ou subeventos. Para a análise de atividades complexas, sistemas de reconhecimento compostos de múltiplas camadas podem ser construídos.

As abordagens de camada única são novamente classificadas em duas categorias, dependendo da forma como modelam as atividades humanas: as *abordagens espaço-temporais*, que interpretam um vídeo de entrada como um volume tridimensional XYT (sendo X e Y os eixos espaciais e T o eixo temporal) e as *abordagens sequenciais,* que interpretam o vídeo como uma sequência de observações.

As abordagens espaço-temporais são, por sua vez, classificadas em três categorias, de acordo com os atributos que são utilizados dos volumes tridimensionais espaço-tempo: os próprios *volumes espaço-tempo*, suas *trajetórias* ou *descritores de ponto de interesse local*. As abordagens sequenciais também são novamente classificadas de acordo com a metodologia que utilizam, sendo divididas em duas novas categorias: *metodologias de reconhecimento baseadas em exemplos,* que representam as atividades humanas mantendo uma sequência ou um conjunto de sequências de amostra de execuções de ações e *metodologias de reconhecimento baseadas em estados*, que representam as atividades humanas como um modelo composto por um conjunto de estados.

As abordagens hierárquicas são classificadas com base nas metodologias de reconhecimento que utilizam, podendo ser *abordagens estatísticas*, *abordagens sintáticas* e *abordagens baseadas em descrição*.

As abordagens estatísticas representam e reconhecem ações humanas de alto nível, através da construção de modelos estatísticos baseados em estados estruturados de forma hierárquica, empregando comumente técnicas como modelos ocultos de Markov (HMMs – em inglês "hidden Markov models") (57) e redes bayesianas dinâmicas (DBNs – em inglês "dynamic bayesian networks") (40). De forma similar, as abordagens sintáticas utilizam uma sintática gramatical para modelar atividades sequenciais, como a gramática estocástica livre de contexto (SCFG – em inglês "stochastic context-free grammar") (26). As abordagens

baseadas em descrição representam as ações humanas através da descrição dos subeventos que compõem as atividades e suas estruturas lógicas, espaciais e temporais (1; 28).

Na próxima seção serão apresentados trabalhos relacionados, que também utilizaram técnicas de visão computacional e inteligência artificial no domínio da patinação artística, visando a avaliação da qualidade de ações.

3.2 TRABALHOS RELACIONADOS

Uma das áreas abordadas neste trabalho (visão computacional para tarefas de avaliação das ações), apesar de ainda pouco explorada em termos de pesquisas e trabalhos já desenvolvidos, apresenta grande potencial de crescimento e aplicação. Assim como em diversos outros aspectos da sociedade (educação, formas de trabalho, rotina diária das pessoas etc.), a aplicação da tecnologia, bem como a expectativa das pessoas por novas formas de interação, está em forte expansão para aspectos ligados ao esporte e às artes.

Diversos trabalhos na área de visão computacional e inteligência artificial para reconhecimento e avaliação de ações humanas foram analisados, para maior entendimento do tema e das técnicas comumente aplicadas. Alguns deles se destacam por abordarem temáticas e/ou técnicas semelhantes às que serão utilizadas neste trabalho, como tarefas de pré-processamento de imagens e vídeos, uso de redes neurais profundas e aprendizado de máquina.

Os trabalhos apresentados nas seções 3.2.1, 3.2.2 e 3.2.3 abordam a temática da avaliação da qualidade das ações por meio do uso de estruturas com redes profundas, verificando e avaliando a efetividade das arquiteturas propostas nos domínios do esporte e da patinação artística, relacionando-se com a proposta dessa dissertação por tratarem do mesmo problema e domínio, porém com diferentes abordagens quanto ao problema e à utilização das técnicas de aprendizado de máquina. Cada uma das seções dedica-se exclusivamente à apresentação de um dos trabalhos relacionados selecionados.

3.2.1 “Avaliando a Qualidade das Ações” - Pirsiavash, Vondrick e Torralba (42)

Pirsiavash, Vondrick e Torralba (42) apresentam uma proposta focada na avaliação da qualidade de execução das ações, e não no reconhecimento das ações em si. Os autores apresentam a necessidade da evolução de pesquisas com esse foco. Enquanto a questão do reconhecimento de ações na área de visão computacional teve muitos avanços recentemente,

disponibilizando atualmente algoritmos e metodologias confiáveis para a execução desta tarefa, o problema de avaliar o quão bem as ações são executadas ainda permanece bastante inexplorado na área de visão computacional. Alguns poucos trabalhos anteriores apresentaram propostas promissoras para esse problema, no entanto as soluções apresentadas até o momento são fortemente voltadas e desenvolvidas para ações específicas.

No domínio dos esportes, a avaliação da qualidade de ações poderia permitir que um atleta realizasse seu treino em frente à uma câmera, obtendo em tempo real as respectivas pontuações para os elementos executados, fornecendo ao atleta a oportunidade de aprimorar sua performance. Avaliadores humanos como treinadores, jurados e especialistas normalmente são treinados por muitos anos para desenvolver regras e métodos complexos para avaliar a qualidade das ações. Essa questão é beneficiada no domínio dos esportes, já que muitas imagens de transmissão das competições são disponibilizadas online gratuitamente, junto com rigorosa avaliação e anotação das pontuações realizada pelos jurados especializados das competições (42).

Avaliando processos para reconhecimento de ações, Pirsiavash, Vondrick e Torralba (42) observaram que muitos trabalhos utilizam “bag-of-words”, pontos de interesse, aprendizagem de atributos e abordagens baseadas em poses humanas. Apesar destes atributos terem sido utilizados com o foco de reconhecimento de ações, os autores acreditam que alguns deles, principalmente a aprendizagem de atributos e as abordagens baseadas em poses humanas, podem também ser utilizados para avaliação das ações, com a realização de pequenos ajustes. Assim, para o aprendizado de um modelo de regressão para a avaliação das ações, foram extraídos atributos espaço-temporais dos vídeos. Dois conjuntos de atributos foram considerados, sendo eles atributos de baixo nível (capturam gradientes e velocidades diretamente dos pixels) e atributos de alto nível (baseados na trajetória da pose humana). Através do cálculo do gradiente máximo, é possível encontrar qual articulação o atleta deve mover e em qual direção, a fim de obter a melhor possibilidade de aperfeiçoamento da pontuação.

A proposta apresentada em (42) também possuía como objetivo encontrar quais segmentos dos vídeos contribuíam mais para a qualidade da ação em termos de pontuação. Essa abordagem é extremamente útil para propor formas de resumir os vídeos, encontrando os segmentos de ações que contribuem para pontuações altas ou baixas. Todos os vídeos do conjunto de dados elaborado foram anotados com o quadro inicial e final de cada elemento, além de ser realizada a extração das pontuações oficiais pelos jurados.

Algumas importantes observações e questões foram levantadas pelos autores em (42). O domínio da patinação artística mostrou-se mais desafiador do que outros esportes estudados, já que a variação nas imagens é maior em termos de poses humanas, roupas e áreas de oclusões. Essa observação sugere a necessidade de melhorias na construção do conjunto de dados, tanto pelo aprimoramento de técnicas quanto pelo possível uso de sensores de profundidade. Também é discutida a proposta de se utilizar um vídeo modelo, com uma execução perfeita do elemento, e avaliar a qualidade da ação executada por meio da comparação e verificação das diferenças em relação ao modelo. Este seria um método bastante simples de ser aplicado, porém na prática é extremamente difícil estabelecer este modelo, já que diversos elementos podem ter várias execuções ideais.

Adicionalmente aponta-se que, apesar da avaliação da qualidade de ação ser uma tarefa subjetiva, os jurados das competições conseguem realizar correlações, o que implica a existência de regras bem definidas, de forma que um sistema de visão computacional também deveria poder aprender com os dados. Por fim, é observado que o método proposto, ainda que promissor, é um passo inicial para as pesquisas e avanço desta problemática na área de visão computacional. Ainda há um longo caminho necessário para que os sistemas consigam ter performances semelhantes às de jurados especializados, indicando a grande oportunidade para a realização de novas pesquisas, a fim de aprimorar esta difícil e importante tarefa (42).

3.2.2 “Aprendendo a pontuação em eventos olímpicos” - Parmar e Morris (41)

Parmar e Morris (41) apresentam um estudo focado na avaliação da qualidade de ações em esportes olímpicos. Observam que, enquanto o tópico de reconhecimento de ações foi extensivamente abordado e pesquisado, a questão da avaliação da qualidade de ações permanece ainda pouco explorada e, nas poucas abordagens apresentadas, as propostas são limitadas e aplicáveis apenas para cenários específicos.

Dois importantes fatores são citados como dificuldades enfrentadas neste tópico. Enquanto no reconhecimento de ações existe, normalmente, uma significante diferença entre duas classes distintas de ações, no caso da avaliação da qualidade de ações as diferenças podem ser muito sutis. Outro fator é que, muitas vezes, uma ação pode ser reconhecida ainda que observada de forma parcial, o que não é válido para a avaliação de ações, já que não faz sentido medirmos a qualidade de uma ação observando apenas uma pequena parte de sua execução (41).

Os autores observam que a técnica das redes neurais convolucionais, especialmente a rede neural 3D (C3D), está sendo cada vez mais utilizada em trabalhos para reconhecimento de ações. No entanto, enfatizam que o treinamento de redes profundas demanda grandes conjuntos de dados, sendo este um fator desafiador para a aplicação das técnicas na avaliação da qualidade de ações. Em comparação com o reconhecimento de ações, a quantidade de conjuntos de dados disponíveis para avaliação de ações é bem pequena, existindo ainda a dificuldade de aumentar os conjuntos de dados disponíveis. No tema de reconhecimento de ações há a possibilidade de se utilizar um script que busque novas amostras em sites de vídeos, como o YouTube. No tema de avaliação de ações isso não pode ser diretamente aplicado, já que os dados requerem anotações qualificadas quanto à pontuação das ações realizadas, o que torna a tarefa bem mais trabalhosa (41).

A proposta abordada em (41) visa avaliar tanto a pontuação dos componentes de uma ação, quanto o resultado da ação como um todo. Propostas anteriores preveem apenas a pontuação final obtida, sem realizar diferenças, por exemplo, entre pontuações de execução e de dificuldade. Em algumas outras abordagens a questão é tratada como um problema de classificação, no qual o objetivo é identificar se uma dada execução de um elemento apresentou erros, gerando no final apenas um indicador de “bom” (sem erros) ou “ruim” (apresentou erros) para a execução avaliada, mas não citando o quão bom ou ruim foi a execução.

Em (41), três estruturas são propostas para a avaliação da qualidade das ações, com base em uma versão menor da rede C3D. Na primeira estrutura, uma máquina de vetor de suporte (SVR – em inglês “support vector machine”) é implementada diretamente nos atributos da C3D. A segunda estrutura modela claramente a natureza sequencial de uma ação, usando uma rede de memória de longo prazo (LSTM – em inglês “long short-term memory”). A terceira estrutura, por fim, combina a LSTM com a SVR. Algumas características importantes são observadas no desenvolvimento dessas propostas:

- as LSTM’s possuem célula de memória, o que facilita o armazenamento e a saída de informações, permitindo assim explorar melhor as relações temporais;
- a rede C3D possui como vantagem o fornecimento de uma descrição compacta para 16 quadros de um vídeo. No uso de redes 2D, cada quadro teria uma descrição individual, o que aumenta o tempo de execução do sistema;
- normalmente, sequências mais longas de quadros necessitam de mais camadas LSTM para modelar a evolução temporal, enquanto sequências mais curtas podem ser modeladas com apenas uma camada;

- para realizar a avaliação tanto da pontuação dos componentes de uma ação, quanto do resultado da ação como um todo, algumas questões adicionais são abordadas. Para o modo de treinamento supervisionado, é necessária a identificação e anotações de rótulos para cada subação, assim como para as pontuações acumuladas. No entanto, em um clipe de 16 quadros de uma rede C3D, não é garantido que uma subação será coberta de forma completa, gerando a necessidade de algum outro método para atribuir com precisão as pontuações para as subações, o que demanda alto conhecimento de aplicação e implementação, além esforços substanciais na rotulagem. Neste caso, o modo de treinamento sem supervisão pode economizar tempo e esforço, fornecendo uma solução escalável.

As três estruturas propostas em (41) diferem principalmente na maneira que agregam atributos da C3D no nível dos clipes, para a obtenção da descrição no nível do vídeo, que é expressiva sobre a qualidade da ação. A estrutura C3D-SVR apresentou melhores resultados, porém não foi capaz de detectar erros cometidos no decorrer da execução de uma ação. O desempenho da estrutura C3D-LSTM foi melhorado com a utilização da SVR e, embora os resultados sejam inferiores aos obtidos com a C3D-SVR, existe nessa estrutura a vantagem de poderem ser identificados os segmentos com erros na execução de uma ação. No domínio da patinação artística, os resultados obtidos foram muito fracos, apresentando em torno de 0,22 de correlação. O baixo desempenho foi atribuído ao uso do conjunto de dados UCF-101 para o pré-treinamento das redes, já que o conjunto não apresenta imagens de patinação artística e, portanto, as amostras apresentadas da patinação não foram suficientemente expressivas.

3.2.3 “Aprendendo a pontuação de vídeos do esporte patinação artística” - Xu et al. (54)

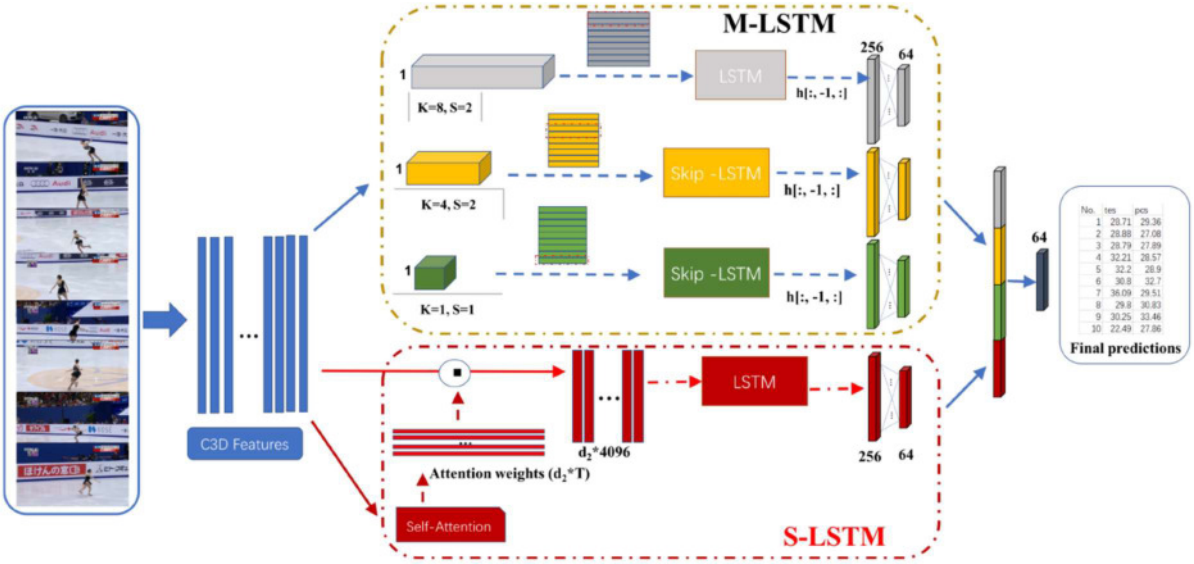
Xu et al. (54), assim como outros autores que abordaram o tópico da avaliação da qualidade das ações, observam que este tema ainda foi pouco abordado pela comunidade científica, especialmente quando comparado com o tema de reconhecimento das ações. Complementam que, no domínio da patinação artística, pouquíssimos trabalhos foram publicados, citando apenas os sistemas propostos por Pirsiavash, Vondrick e Torralba (42) e Parmar e Morris (41), apresentados nas seções anteriores desta dissertação. Os principais desafios apontados para o pequeno número de trabalhos na área foram a longa duração dos vídeos com performances de patinação artística (em média 2 minutos e 50 segundos), a necessidade de conhecimento altamente especializado para a rotulagem dos vídeos e o fato de que nem todos os segmentos da apresentação são úteis para a avaliação das pontuações, já que

os jurados avaliam principalmente a execução dos elementos técnicos (TES) e um conjunto de habilidades artísticas (PCS).

As pesquisas recentes nesta área de visão computacional estão utilizando, frequentemente, arquiteturas profundas para extrair representações de atributos de forma efetiva no domínio dos vídeos. Técnicas como redes convolucionais (ConvNets – em inglês "convolutional networks"), memórias de longo prazo (LSTMs) e redes convolucionais tridimensionais (C3D) são amplamente aplicadas para o aprendizado de atributos espaço-temporais (54).

No trabalho apresentado em (54), os autores contribuem de três formas principais: através da criação de um conjunto de dados de vídeos de patinação artística nomeado FIS-V, por meio da proposta de dois modelos para as atividades de regressão - nomeados "Self-Attentive LSTM" (S-LSTM) e "Multi-scale Convolutional Skip LSTM" (M-LSTM) e, por fim, propondo uma estrutura completa que integra dois modelos de rede propostos. Uma visão geral desta estrutura pode ser verificada na figura 17.

Figura 17 – Visão geral do modelo de agregação convolucional multiescalas com skip-LSTM



Fonte: "adaptado de" Xu et al. (54)

Os autores avaliam os modelos propostos em dois conjuntos de dados, sendo eles o MITskate (42) e o conjunto próprio FIS-V. O conjunto por eles criado é composto de 500 vídeos de apresentações do programa curto individual feminino e foi cuidadosamente montado e pré-processado. Os vídeos possuem anotações para o nome do atleta e da competição, além de rótulos para as pontuações técnicas (TES) e de performance (PCS), que

foram obtidas oficialmente pelos jurados de avaliação da competição. O conjunto foi aleatoriamente dividido em 400 vídeos de treinamento e 100 vídeos de teste (54).

Para o processamento dos vídeos em (54), redes convolucionais espaço-temporais profundas foram utilizadas, de forma a se obter uma melhor representação dos atributos das imagens. As redes convolucionais 3D foram pré-treinadas no conjunto de dados de larga escala "Sports-1M" (29), que contém 1.133.158 vídeos anotados automaticamente com rótulos de 487 esportes. Assim como na proposta apresentada por Parmar e Morris (41), foram utilizadas janelas de 16 quadros para a avaliação e descrição dos vídeos. Os modelos de rede propostos (S-LSTM e M-LSTM) conseguem lidar com o problema dos vídeos de patinação artística serem muito longos, o que não consegue ser processado por um modelo LSTM comum.

Os resultados encontrados no trabalho (54) são bastante promissores, especialmente quando comparados aos resultados obtidos em propostas anteriores, como (42) e (41). O desempenho da estrutura proposta apresentou o melhor desempenho em ambos conjuntos de dados, superando por uma grande margem as propostas anteriores, comprovando a eficácia da proposta.

Estudos também foram conduzidos para verificar a contribuição individual dos modelos S-LSTM e M-LSTM propostos. Os resultados individuais do M-LSTM já puderam superar todas as outras linhas de base, em ambos conjuntos de dados avaliados. Estes resultados são considerados razoáveis, já que o modelo proposto M-LSTM consegue aprender efetivamente informações locais e globais. No conjunto de dados MITskate, o modelo S-LSTM é complementar ao M-LSTM, proporcionando a obtenção de resultados melhores. No conjunto de dados FIS-V, ambos modelos M-LSTM e S-LSTM apresentaram desempenho muito bom (54).

Adicionalmente, os autores em (54) realizaram uma comparação no desempenho da estrutura proposta ao avaliar as pontuações técnica (TES) e de performance (PCS). Apesar do caráter mais subjetivo da avaliação PCS, os resultados obtidos foram geralmente melhores do que para a avaliação dos elementos técnicos, o que reflete que a rede conseguiu prever de maneira relativamente mais fácil os critérios dos elementos de apresentação.

4 MÉTODO E IMPLEMENTAÇÃO

Neste capítulo são apresentadas em detalhes todas as etapas realizadas no desenvolvimento da proposta deste trabalho, bem como as técnicas e recursos utilizados. Inicialmente é apresentado um diagrama esquemático para entendimento da abordagem do problema e, na sequência, são apresentados os modelos propostos e como eles atenderam o objetivo de classificar e avaliar automaticamente performances de elementos técnicos da patinação artística.

4.1 ABORDAGEM DO PROBLEMA E DETALHAMENTO DA PROPOSTA

Este trabalho tem como objetivo principal desenvolver e apresentar os resultados de uma proposta para identificação e avaliação automática da execução de elementos técnicos da patinação artística. Através de diversas etapas que aplicam técnicas de visão computacional e aprendizado de máquina, os modelos propostos são capazes de observar diferentes execuções de elementos técnicos, entender qual é o movimento apresentado e avaliar o quão bem o elemento foi executado.

A abordagem para o problema apresentado se baseia em classificação de imagens, através do treinamento supervisionado de modelos de classificadores. O processo de treinamento dos modelos é realizado através de imagens no formato “quadro a quadro” (em inglês “frame-by-frame” ou “stop motion”), que apresentam em uma só imagem a execução completa do elemento técnico apresentado, montadas a partir de vídeos de competições oficiais de patinação artística. Os atributos destas imagens foram extraídos por meio de uma rede neural convolucional pré-treinada e posteriormente utilizados para a finalização do treinamento através de outras técnicas de aprendizado de máquina, já que o conjunto de dados é pequeno para uma abordagem completa do problema através da aprendizagem profunda. Diferentes métodos de classificação foram avaliados no treinamento, a fim de possibilitar a comparação da acurácia dos diferentes classificadores e a definição dos que se adequam melhor à proposta apresentada. Um diagrama esquemático desta proposta é apresentado na figura 18.

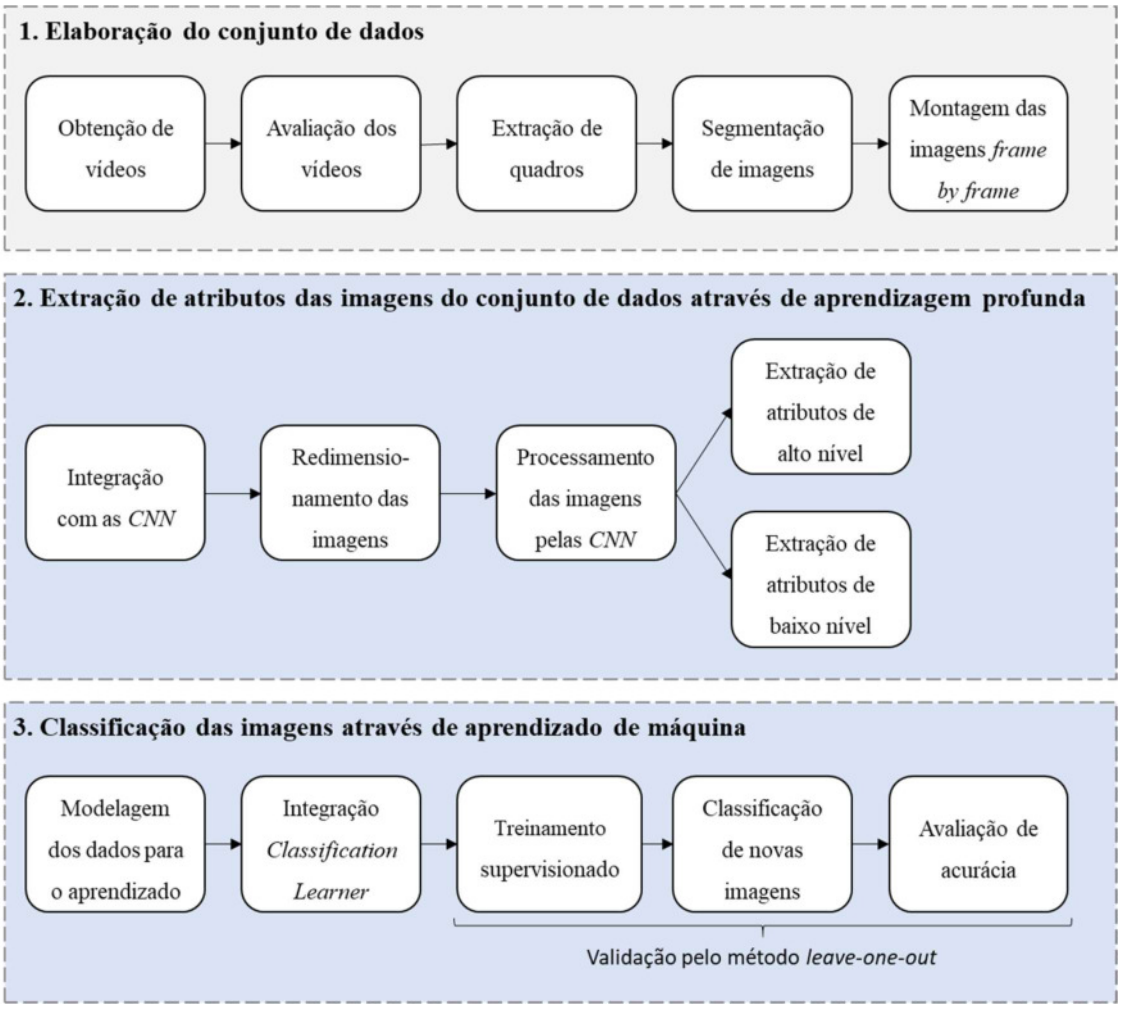
Conforme o diagrama apresentado na figura 18, o desenvolvimento deste trabalho se deu em três etapas principais:

Etapa 1 - Elaboração do conjunto de dados;

Etapa 2 - Extração de atributos das imagens do conjunto de dados através de aprendizagem profunda;

Etapa 3 - Classificação das imagens através de aprendizado de máquina.

Figura 18 – Diagrama esquemático da proposta apresentada nesta dissertação



Fonte: Autora.

A etapa 1 (destacada na cor cinza da figura 18) foi realizada por um agente humano com auxílio de ferramentas computacionais para processamento de vídeos e imagens. As etapas 2 e 3 (destacadas na cor azul da figura 18) foram realizadas por meio de modelos artificiais, através de uma combinação de técnicas de aprendizado de máquina. Os modelos foram em Matlab, nas versões R2018b e R2019a, com os seguintes complementos:

- *Image Processing Toolbox* – versão 10.4;
- *Computer Vision Toolbox* – versão 9.0;

- *Deep Learning Toolbox* – versão 12.1;
- *Deep Learning Toolbox models for ResNet-18 and Xception networks* – versão 19.1.0;
- *Parallel Computing Toolbox* – versão 7.0;
- *Statistics and Machine Learning Toolbox* – versão 11.5.

As seções 4.2, 4.3 e 4.4 apresentam a descrição detalhada das atividades realizadas nas etapas 1, 2 e 3 (figura 18), respectivamente.

4.2 ELABORAÇÃO DO CONJUNTO DE DADOS

Para o desenvolvimento deste trabalho foram verificados diversos conjuntos de dados já existentes e atualmente disponíveis de forma aberta na internet, porém não foi possível encontrar nenhum que apresentasse imagens com a execução de elementos da patinação artística, conforme desejado neste trabalho. Alguns trabalhos relacionados desenvolveram conjuntos com dados de patinação artística, porém os conjuntos eram compostos por vídeos. A possibilidade de utilizar o conjunto de vídeos criado em (54), como base para a extração de quadros e montagem das imagens, foi uma das possibilidades avaliadas. No entanto, o conjunto não foi encontrado disponível na internet durante a realização desse trabalho, levando à criação de um conjunto totalmente novo e específico.

Nesta proposta, foi escolhido trabalhar com um conjunto de dados de imagens ao invés de vídeos, por duas razões principais:

1. diminuir a complexidade do conjunto de dados, da extração de atributos e do treinamento dos modelos classificadores;
2. buscar utilizar recursos tecnológicos que sejam amplamente aceitos no domínio dos esportes e da patinação artística, como o uso da composição de imagens quadro a quadro para a descrição de movimentos, de forma a tornar a proposta mais próxima da realidade do domínio e utilizando elementos de valor agregado comprovado.

Para a obtenção dos dados iniciais, foram escolhidos vídeos que estão disponíveis de forma aberta na internet. A fim de obtermos dados (vídeos e posteriormente imagens) com maior qualidade e de termos pontuações oficiais já determinadas por equipes especializadas, apenas vídeos de competições oficiais de patinação artística, em resolução full HD – 1080p foram utilizados. Junto com os vídeos, foram obtidos no website da União Internacional de Patinação – ISU (23) os livros de resultados oficiais das competições avaliadas, permitindo assim uma correta rotulagem dos elementos técnicos e das pontuações obtidas.

Os vídeos de competições de patinação artística são bastante desafiadores para a montagem de imagens quadro a quadro e a classificação e avaliação das execuções dos elementos representados nas imagens elaboradas, já que as competições são filmadas através de câmeras em plataformas móveis, com os atletas executando os elementos em diferentes posicionamentos em relação à pista de patinação e às câmeras. Dessa forma, um mesmo elemento pode ser capturado "de frente", "de lado", ou "de costas", dependendo da posição de patinação do atleta e das câmeras naquele dado momento. Essa característica constitui um desafio adicional tanto para a avaliação manual dos vídeos, exigindo conhecimento dos saltos apresentados e constante consulta ao livro de resultados oficiais, quanto para os algoritmos classificadores, que devem ser capazes de reconhecer as diferentes referências espaciais para o mesmo tipo de elemento apresentado.

Dois elementos técnicos foram escolhidos para a avaliação desta proposta, sendo eles os saltos *Triple Axel* (Axel com três rotações) e *Quad Toe* (toe loop com quatro rotações). Estes elementos foram escolhidos por serem frequentemente apresentados nas competições oficiais, permitindo a obtenção de maior número de dados através os vídeos avaliados. Os vídeos foram manualmente avaliados para a identificação dos quadros que continham apresentações dos elementos técnicos escolhidos e, com o auxílio do software aberto de edição de vídeos *GOM Player*, foram extraídas as imagens quadro a quadro que compunham a execução completa dos elementos. Os quadros foram extraídos em uma taxa de 2 qps (quadros por segundo), já que os movimentos possuem execução extremamente rápida, gerando em torno de 10 a 12 quadros para a representação completa de cada elemento. No total foram extraídos aproximadamente 1000 quadros, para representar 90 diferentes apresentações dos elementos técnicos selecionados.

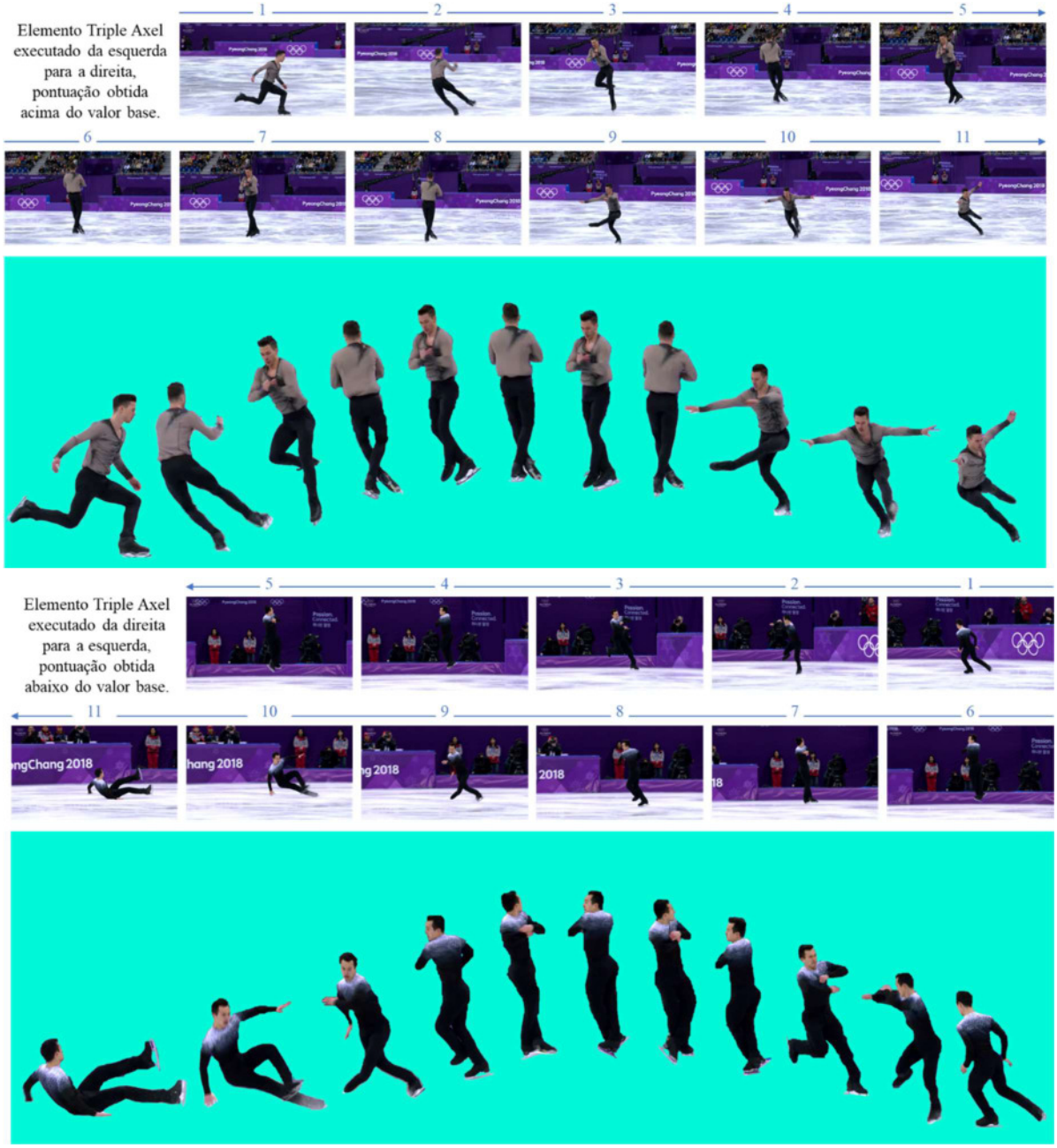
Após a extração dos quadros, todas as imagens foram segmentadas individualmente com o auxílio da função *Image Segmenter* do Matlab, que realiza a segmentação de imagens por refinamento de regiões, através do algoritmo *graph-cut* interativo. Em todas as imagens, os atletas foram extraídos do cenário, de forma a levar para a próxima etapa apenas as informações referentes aos seus movimentos e posições corporais, eliminando informações referentes ao fundo como arquibancada, público, jurados, equipes técnicas e pista de patinação, que apenas trariam maior complexidade para as futuras tarefas de classificação de imagens, aparentemente não agregando valor aos resultados.

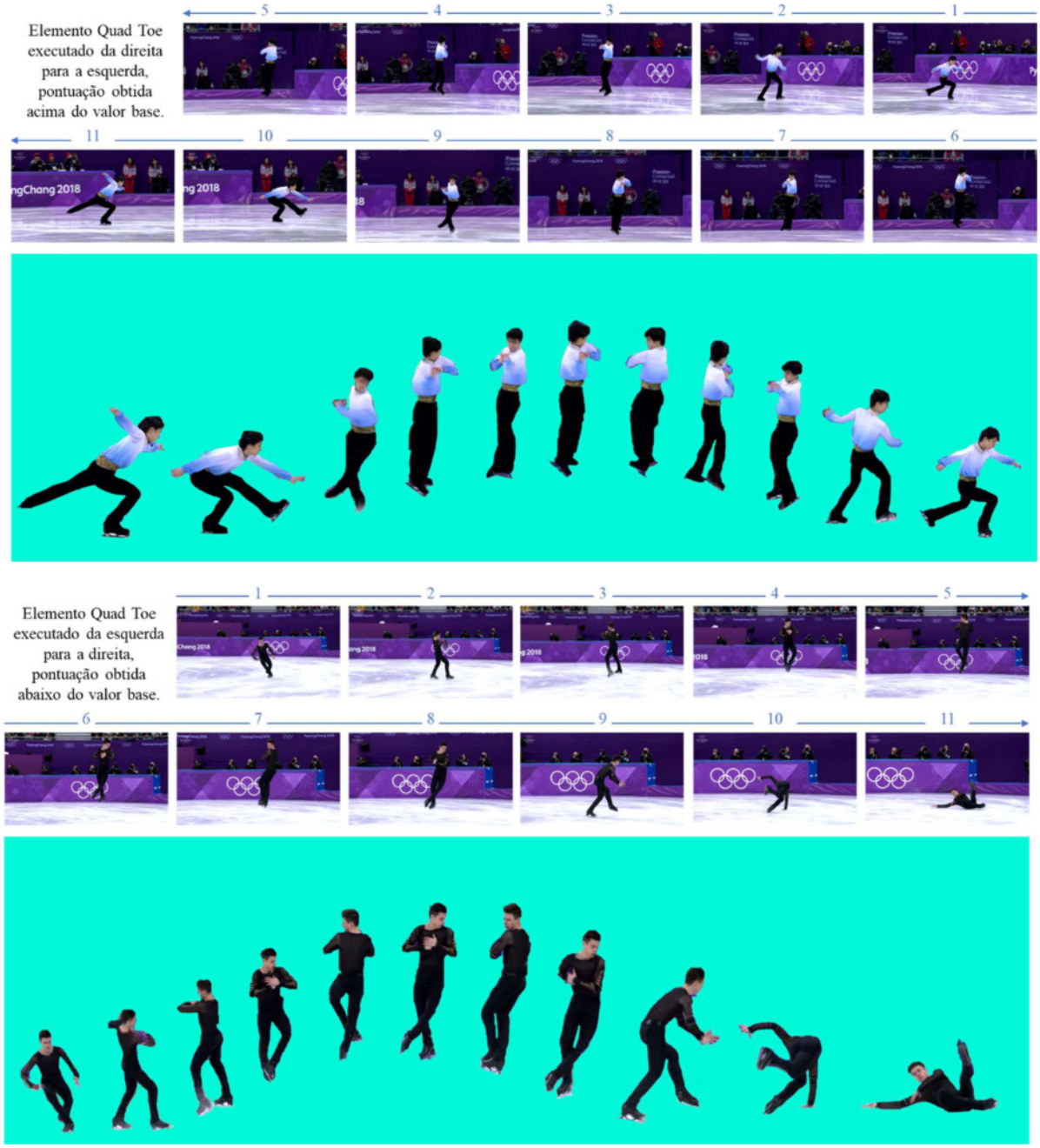
A partir das imagens segmentadas dos quadros que compõem a execução dos movimentos, foram criadas novas imagens contendo, cada uma, uma execução completa do

elemento. As imagens foram montadas com o auxílio do software aberto de edição de imagens GIMP, possuindo 5.000 x 1.500 pixels, 72 ppp e cor RGB.

O conjunto de dados elaborado possui 90 imagens, sendo 48 de execuções do elemento técnico *Triple Axel* (27 com pontuações obtidas acima do valor base e 21 com pontuações obtidas abaixo do valor base) e 42 de execuções do elemento técnico *Quad Toe* (27 com pontuações obtidas acima do valor base e 15 com pontuações obtidas abaixo do valor base). A figura 19 apresenta exemplos de imagens geradas, com a composição quadro a quadro dos movimentos executados. Nesta imagem é possível observar alguns exemplos dos diferentes posicionamentos dos atletas em relação às câmeras.

Figura 19 – Exemplos de imagens quadro a quadro montadas para o conjunto de dados





Fonte: Autora.

4.3 EXTRAÇÃO DE ATRIBUTOS DAS IMAGENS ATRAVÉS DE APRENDIZAGEM PROFUNDA

As técnicas avaliadas nesta proposta para o aprendizado de máquina demandam a descrição das imagens através de seus atributos, a serem utilizados como dados de entrada. A escolha dos atributos a serem utilizados é de extrema importância e pode afetar diretamente os resultados, sendo então necessária a escolha dos atributos mais relevantes para que o conjunto de dados seja representado de forma efetiva. Os atributos de uma imagem podem ser

considerados de baixo ou alto nível. Atributos de baixo nível são os que descrevem a imagem em um nível de características mais primitivas, como cor, pixel, forma, textura, orientação de gradientes, entre outros. Atributos de alto nível são os que descrevem a imagem em um nível de contexto e conhecimento, apresentando características mais abstratas, ou de semântica, sobre o conteúdo da imagem.

O conjunto de dados elaborado neste trabalho é bastante pequeno para tarefas de treinamento de rede, conforme apresentado na seção anterior, o que fortalece ainda mais a necessidade de uma escolha efetiva dos atributos das imagens. A fim de obter atributos que sejam realmente representativos para as atividades de classificação, a extração dos atributos foi realizada por meio de redes neurais profundas, pré-treinadas em grandes conjuntos de imagens. Duas redes foram escolhidas nesta proposta, de forma a permitir uma melhor avaliação do problema e a comparação da performance de ambas as redes na extração de atributos: ResNet e Xception.

A *ResNet* (21) é uma rede neural convolucional treinada em mais de um milhão de imagens do banco de dados *ImageNet* (15), com versões de diferentes profundidades e tamanho de entrada da imagem de 224 por 224. Uma de suas características é a habilidade que adquiriu, após o treinamento extensivo, de realizar com qualidade a representação de atributos para uma grande variedade de imagens. O banco de dados no qual a ResNet foi treinada é composto por imagens de diversos tipos e temas, não sendo específico para o domínio dos esportes e nem da patinação artística. No entanto, como a rede profunda pré-treinada está sendo utilizada nesta proposta apenas para a extração de atributos de imagens, e não para o treinamento do agente artificial em si, o uso desta rede indica uma boa possibilidade de extração de atributos efetivos para a representação do novo conjunto de dados.

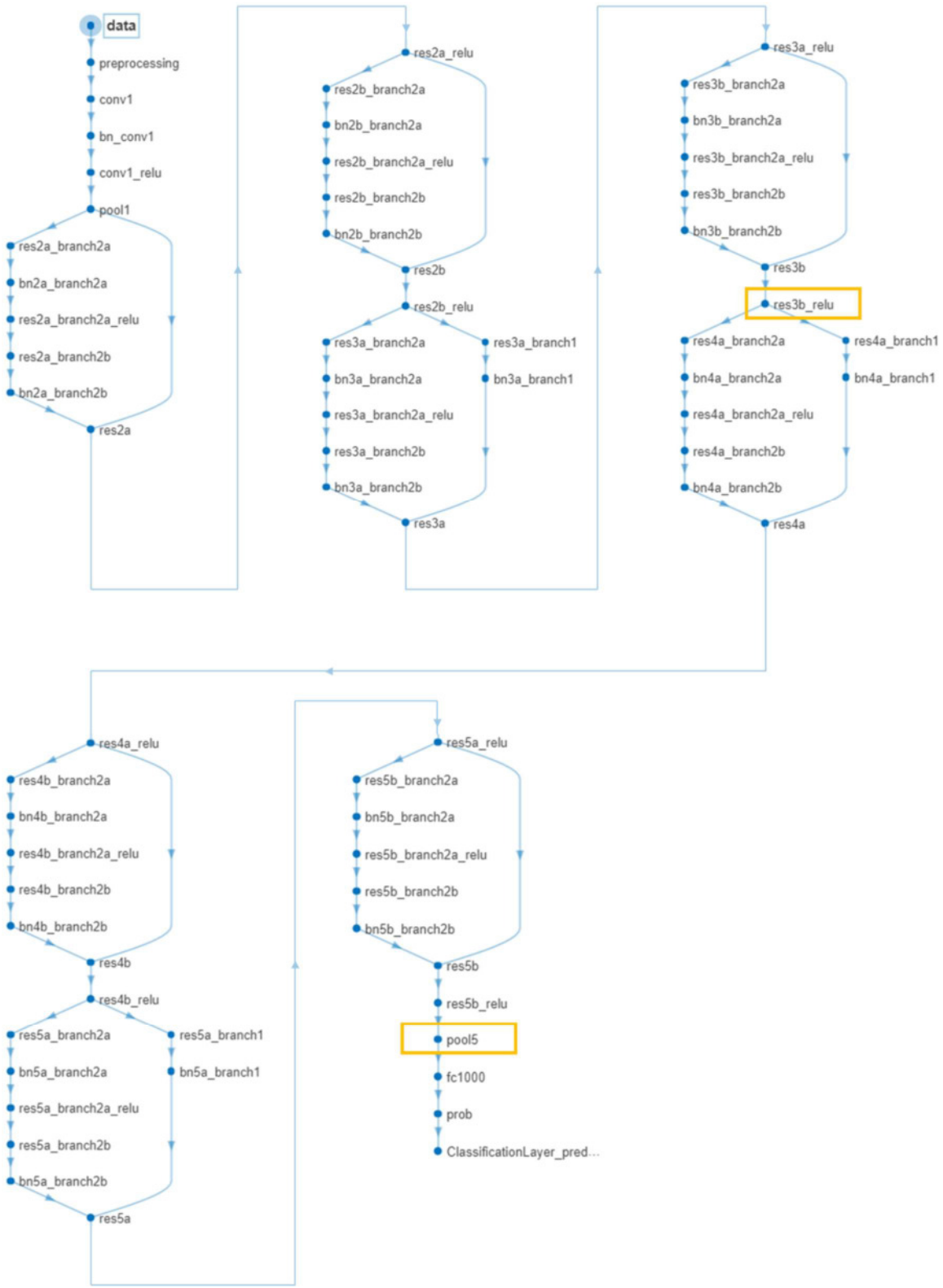
Os atributos de imagem de baixo nível podem ser extraídos nas camadas iniciais da rede, enquanto nas camadas finais da rede encontram-se os atributos de alto nível para a descrição da imagem. Como as imagens a serem avaliadas neste trabalho (imagens quadro a quadro da execução de elementos da patinação artística) são bastante diferentes das imagens utilizadas no pré-treinamento da ResNet (banco de dados *ImageNet)*, acredita-se que os atributos de baixo nível sejam mais adequados para a proposta. No entanto, para efeitos de comparação de resultados nesta abordagem específica, foram extraídos atributos de baixo e alto nível.

Os atributos foram extraídos da rede ResNet através do código descrito na sequência:

```
%Inicializar conjunto de dados
imds = imageDatastore('D:\Banco de
Dados_Mestrado)','IncludeSubfolders',true,'LabelSource','foldernames');

%Carregar rede pré-treinada ResNet
net = resnet18;

%Visualizar a arquitetura da rede
inputSize = net.Layers(1).InputSize;
analyzeNetwork(net)

%Redimensionar as imagens do conjunto de dados para entrada na ResNet
resizedimds = augmentedImageDatastore([224 224 3],imds;

%Extrair atributos de baixo nível da ResNet por meio de ativações na camada
'res3b_relu'
%layer = 'res3b_relu';
%features = activations(net,resizedimds,layer,'OutputAs','rows');
%whos features

%Extrair atributos de alto nível da ResNet por meio de ativações na camada
'pool5'
layer = 'pool5';
features = activations(net,resizedimds,layer,'OutputAs','rows');
whos features

%Ajustar os atributos para o formato I x A, onde N é o número de imagens e
A é o número de atributos, removendo as dimensões de comprimento 1
(singleton) e transpondo
features = squeeze(mean(features,[1 2]))';
whos features

%Extrair os rótulos de classe do conjunto de dados
imdslabels = imds.Labels
```

Através da extração de atributos pela rede ResNet, foram obtidos 128 atributos de baixo nível (camada res3b_relu) e 512 atributos de alto nível (camada pool5) para a descrição de cada uma das imagens do conjunto de dados. A arquitetura do modelo utilizado da rede ResNet para Matlab, bem como as camadas das quais foram extraídos os atributos, podem ser verificadas na figura 20.

Figura 20 – Arquitetura do modelo utilizado da rede ResNet-18 para Matlab, com 72 camadas



Fonte: Autora; imagem gerada pela função AnalyzeNetwork para ResNet-18 model – Matlab.

A *Xception* (10) é também uma rede neural convolucional treinada no bando de dados ImageNet (15), com versões de diferentes profundidades e tamanho de entrada da imagem de 299 por 299. A rede apresenta uma arquitetura desenvolvida mais recentemente, diferente da ResNet, sendo uma rede profunda com camadas de convolução separáveis. Assim como realizado com a rede ResNet, a Xception foi utilizada para a extração de atributos das imagens que compõem o conjunto de dados proposto, tanto de baixo nível quanto de alto nível.

Os atributos foram extraídos da rede Xception através do código descrito na sequência:

```
%Inicializar conjunto de dados
imds = imageDatastore('D:\Banco de
Dados_Mestrado)','IncludeSubfolders',true,'LabelSource','foldernames');

%Carregar rede pré-treinada Xception
%net = xception;

%Visualizar a arquitetura da rede
inputSize = net.Layers(1).InputSize;
analyzeNetwork(net)

%Redimensionar as imagens do conjunto de dados para entrada na Xception
%resizedimds = augmentedImageDatastore([299 299 3],imds;

%Extrair atributos de baixo nível da Xception por meio de ativações na
camada 'add_5'
layer = 'add_5';
features = activations(net,resizedimds,layer,'OutputAs','rows');
whos features

%Extrair atributos de alto nível da Xception por meio de ativações na
camada 'avg_pool'
layer = 'avg_pool';
features = activations(net,resizedimds,layer,'OutputAs','rows');
whos features

%Ajustar os atributos para o formato I x A, onde N é o número de imagens e
A é o número de atributos, removendo as dimensões de comprimento 1
(singleton) e transpondo
features = squeeze(mean(features,[1 2]))';
whos features

%Extrair os rótulos de classe do conjunto de dados
imdslabels = imds.Labels
```

Através da extração de atributos pela rede Xception, foram obtidos 728 atributos de baixo nível (camada add_5) e 2.048 atributos de alto nível (camada avg_pool) para a descrição de cada uma das imagens do conjunto de dados. A arquitetura do modelo utilizado da rede Xception para Matlab, bem como as camadas das quais foram extraídos os atributos, podem ser verificadas na figura 21.

Figura 21 – Arquitetura do modelo utilizado da rede Xception para Matlab, com 171 camadas

Fonte: Autora; imagem gerada pela função AnalyzeNetwork para Xception model – Matlab.

Para o processamento e extração de atributos das imagens através das redes ResNet e Xception, o agente artificial realizou redimensionamento automático dos dados do conjunto por meio da função *augmentedImageDatastore*, de forma a adequar as imagens aos padrões de entrada de cada rede. Após o processamento de todas as imagens do conjunto através das redes profundas, os atributos gerados foram salvos em tabelas, a fim de serem utilizados na sequência para a etapa de treinamento dos modelos classificadores através de aprendizado de máquina.

4.4 CLASSIFICAÇÃO DAS IMAGENS ATRAVÉS DE APRENDIZADO DE MÁQUINA

Após a geração das tabelas com os atributos de imagens extraídos pelas redes profundas pré-treinadas ResNet e Xception, conforme descrito anteriormente na seção 4.3, os rótulos das imagens foram adequadamente transferidos para a tabela, de forma a correlacionar os atributos com as categorias de classificação reais e permitir a realização do treinamento supervisionado.

O treinamento foi realizado utilizando diferentes algoritmos de aprendizado de máquina, de forma a possibilitar uma melhor avaliação dos classificadores e a identificação dos algoritmos mais adequados para o problema abordado nessa dissertação. Os algoritmos avaliados foram:

- Árvores de decisão (31);
- Análise discriminante linear (18; 36);
- Regressão logística (49);
- Classificadores Naïve Bayes (31);
- Máquinas de vetores de suporte (SVM) (9; 13);
- Classificadores KNN (k vizinhos mais próximos) (12; 31);
- Classificadores *Ensemble* (boosted trees, bagged trees, subespaço discriminante, subespaço KNN e RUSBoosted trees) (31).

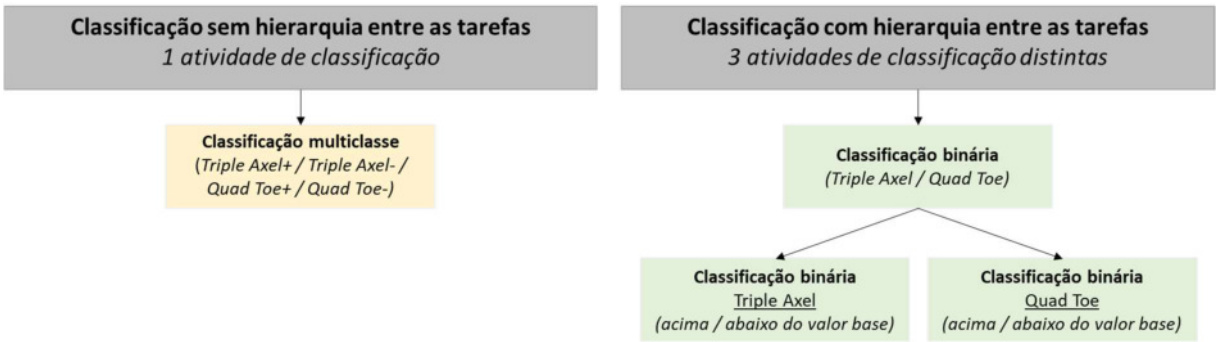
O treinamento foi realizado através do uso do aplicativo *Classification Learner* do Matlab com os parâmetros abaixo citados:

- Uso de todos os atributos extraídos e descritos nas tabelas de dados, sem otimização através de seleção manual ou automática dos atributos mais significativos;
- Análise de componentes principais (função PCA) desativada;

- Avaliação pelo método de validação cruzada *leave-one-out* (30), através da seleção da função de ativação *k-fold* com número máximo de divisões possíveis (número total de elementos do conjunto de dados) para cada situação avaliada;
- Ativação da computação paralela, diminuindo o tempo necessário para o treinamento de todos os classificadores avaliados;
- Os demais parâmetros do *Classification Learner* e dos algoritmos avaliados foram mantidos conforme configuração padrão do Matlab.

Além de realizar o treinamento através de técnicas e algoritmos diferentes, diversas outras investigações foram realizadas, a fim de permitir a comparação de diferentes abordagens e métodos. Todos os algoritmos classificadores foram treinados separadamente tanto com os atributos de baixo nível, quanto com os de alto nível, extraídos através das duas redes profundas avaliadas (ResNet e Xception). As atividades de classificação também foram tratadas de duas formas, considerando em uma delas a hierarquia entre as tarefas e na outra uma classificação direta entre as categorias possíveis, conforme esquema apresentado na figura 22.

Figura 22 – Abordagens utilizadas no treinamento de máquina para as tarefas de classificação



Fonte: Autora.

Na abordagem s*em hierarquia entre as tarefas* as quatro classificações possíveis (Triple Axel+, Triple Axel-, Quad Toe+ e Quad Toe-) foram avaliadas de forma conjunta, tendo os algoritmos classificadores que escolher diretamente entre uma delas. Esta abordagem gerou a necessidade da aplicação de classificações multiclasse, o que aumenta a complexidade da tarefa e demanda uma maior abstração de parâmetros por parte dos modelos.

Na abordagem *com hierarquia entre as tarefas* foi considerada para a realização do aprendizado de máquina a hierarquia existente entre as tarefas de identificação e avaliação do elemento apresentado. Os algoritmos classificadores foram treinados em um primeiro

momento para identificar qual era o elemento sendo executado (Triple Axel ou Quad Toe) e, na sequência, para avaliar a qualidade de execução apresentada (acima ou abaixo do valor base do elemento). Esta abordagem permitiu a aplicação de classificações binárias e mais diretas.

Os resultados obtidos por todos os classificadores avaliados nas diversas abordagens propostas serão apresentados e discutidos no próximo capítulo.

5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Conforme metodologia e etapas de implementação detalhadas no capítulo 4, os modelos classificadores foram treinados através de diferentes técnicas de aprendizado de máquina, a fim de realizar as atividades de identificação e avaliação da execução de elementos técnicos da patinação artística. Nesta seção os resultados obtidos com as diferentes técnicas são apresentados através de tabelas e matrizes de confusão, a fim de permitir a comparação entre as técnicas avaliadas neste problema e um melhor entendimento da performance dos classificadores.

Os resultados serão apresentados em quatro subseções, cada uma delas tratando de uma das diferentes abordagens das tarefas de classificação, conforme explicado anteriormente.

5.1 CLASSIFICAÇÃO SEM HIERARQUIA ENTRE AS TAREFAS

Nesta abordagem existiram quatro categorias possíveis de classificação, sendo elas Triple Axel+ (3A pos), Triple Axel- (3A neg), Quad Toe+ (4T pos) e Quad Toe- (4T neg). Todas as possíveis classes foram treinadas e avaliadas de forma conjunta, transformando o problema em uma classificação multiclasse. A tabela 2 apresenta os valores de acurácia obtidos e as células marcadas em amarelo representam os resultados máximos alcançados.

A partir da avaliação dos resultados apresentados na tabela 2, é possível observar que, na abordagem de classificação multiclasse, nenhuma das técnicas de aprendizado de máquina apresentou um bom desempenho, estando todas com acurácia igual ou menor a 43,3%. O melhor resultado geral foi obtido pela técnica de aprendizado *Cubic KNN*, com o uso dos atributos de baixo nível extraídos pela rede pré-treinada Xception.

Ao compararmos os resultados obtidos pelas duas redes utilizadas para extração de atributos (ResNet e Xception), os valores máximos de acurácia alcançados foram bastante semelhantes, embora alcançados com técnicas de aprendizado diferentes. Realizando uma comparação entre os resultados obtidos com o uso dos atributos de baixo e alto nível extraídos por uma mesma rede, é possível observar que, no geral, os atributos de baixo nível apresentaram resultados um pouco melhores do que os atributos de alto nível. Tal efeito era esperado, já que ambas as redes ResNet e Xception foram pré-treinadas com o conjunto de dados ImageNet, que possui imagens muito diferentes das utilizadas nesse trabalho. Dessa forma, os atributos de alto nível não conseguem representar tão bem as imagens quanto os

atributos de baixo nível, já que as descrevem em um nível mais contextual, e os contextos apresentados nos conjuntos de pré-treinamento e treinamento são bastante diferentes.

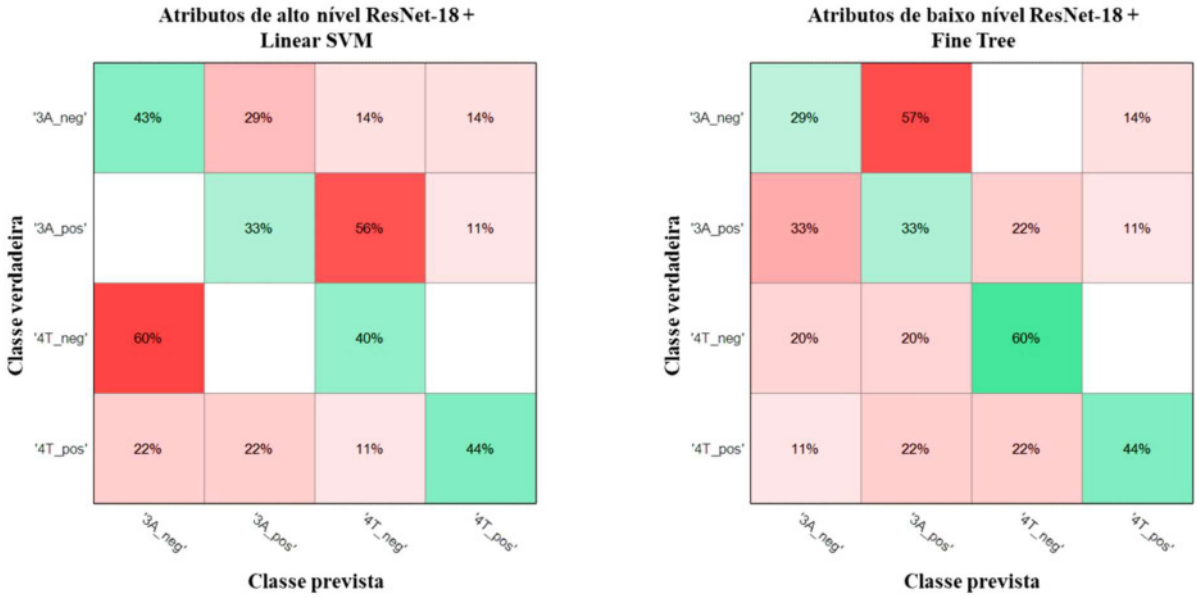
Tabela 2 – Resultados obtidos com a abordagem de classificação multiclasse

	Classificação multiclasse (Triple Axel + / Triple Axel - / Quad Toe + / Quad Toe -)			
	ResNet-18		Xception	
	Atributos de alto nível	Atributos de baixo nível	Atributos de alto nível	Atributos de baixo nível
Fine Tree	13,3%	**40,0%**	10,0%	13,3%
Medium Tree	13,3%	40,0%	10,0%	13,3%
Coarse Tree	13,3%	36,7%	10,0%	13,3%
Linear Discriminant	-	20,0%	-	30,0%
Gaussian Naive Bayes	-	36,7%	-	40,0%
Kernel Naive Bayes	0,0%	36,7%	0,0%	0,0%
Linear SVM	**40,0%**	36,7%	10,0%	30,0%
Quadratic SVM	23,3%	30,0%	20,0%	40,0%
Cubic SVM	36,7%	23,3%	6,7%	20,0%
Fine Gaussian SVM	0,0%	0,0%	0,0%	0,0%
Medium Gaussian SVM	13,3%	26,7%	30,0%	23,3%
Coarse Gaussian SVM	0,0%	0,0%	0,0%	0,0%
Fine KNN	13,3%	23,3%	10,0%	16,7%
Medium KNN	23,3%	20,0%	26,7%	40,0%
Coarse KNN	0,0%	0,0%	0,0%	0,0%
Cosine KNN	13,3%	23,3%	23,3%	20,0%
Cubic KNN	26,7%	30,0%	30,0%	**43,3%**
Weighted KNN	16,7%	33,3%	16,7%	36,7%
Ensemble - Boosted Trees	0,0%	0,0%	0,0%	0,0%
Ensemble - Bagged Trees	23,3%	30,0%	13,3%	20,0%
Ensemble - Subspace Discriminant	20,0%	20,0%	13,3%	26,7%
Ensemble - Subspace KNN	3,3%	13,3%	20,0%	36,7%
Ensemble - RUSBoosted Trees	30,0%	23,3%	**33,3%**	36,7%

Fonte: Autora.
Legenda: Valores de acurácia apresentados por cada uma das técnicas de aprendizado de máquina avaliadas.

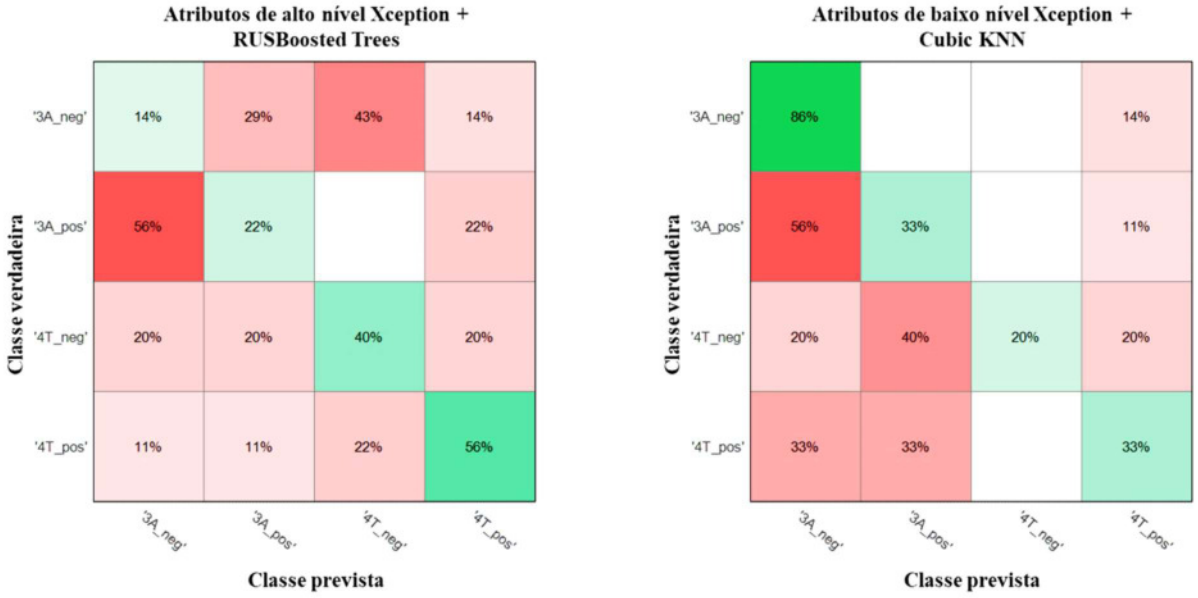
Através da análise das matrizes de confusão é possível entender como foi a performance dos classificadores entre os grupos de classificação existentes. A figura 23 apresenta as matrizes de confusão das técnicas que apresentaram os melhores resultados com os atributos de baixo e alto nível extraídos pela rede ResNet. A figura 24, de forma similar, apresenta as matrizes de confusão das técnicas que apresentaram os melhores resultados com os atributos de baixo e alto nível extraídos pela rede Xception.

Figura 23 – Matrizes de confusão na abordagem de classificação multiclasse, com atributos extraídos pela ResNet



Fonte: Autora.
Legenda: Matrizes de confusão das técnicas de aprendizado de máquina que apresentaram melhores resultados de acurácia nesta abordagem, com a extração de atributos de alto e baixo nível pela ResNet-18.

Figura 24 – Matrizes de confusão na abordagem de classificação multiclasse, com atributos extraídos para Xception



Fonte: Autora.
Legenda: Matrizes de confusão das técnicas de aprendizado de máquina que apresentaram melhores resultados de acurácia nesta abordagem, com a extração de atributos de alto e baixo nível pela Xception.

Na figura 23 é possível observar que para os elementos da classe Triple Axel- (3A neg), a maior confusão com ambos os níveis de atributos e as duas diferentes técnicas foi com a classe Triple Axel+ (3A pos). Ambas classes representam o mesmo elemento técnico da

patinação artística, porém com avaliações da qualidade de execução diferentes (positiva e negativa). Para a técnica Linear SVM com o uso de atributos de alto nível é possível observar que também houve bastante dificuldade de identificação entre os elementos das classes Triple Axel+ (3A pos) e Quad Toe- (4T neg). Para a técnica Fine Tree com o uso de atributos de baixo nível, com exceção da dificuldade de diferenciação já abordada acima entre a classe Triple Axel^{-} e Triple Axel^{+}, as demais classes apresentaram valores falsos positivos distribuídos de forma mais equilibrada entre as possíveis classificações.

Já na figura 24 é possível observar que para os elementos da classe Triple Axel+ (3A pos), a maior confusão com ambos os níveis de atributos e as duas diferentes técnicas foi com a classe Triple Axel- (3A neg), efeito contrário ao obtido com os atribuídos extraídos pela ResNet.

Através da avaliação da tabela e das matrizes de confusão é possível observar a ausência de uma tendência clara de diferenciação tanto entre as técnicas de aprendizado avaliadas, quanto entre as diferentes redes usadas para extração de atributos e pelos atributos de diferentes níveis. Esse comportamento pode estar associado, de forma principal, a dois fatores: não consideração da hierarquia entre as tarefas de classificação e o tamanho do conjunto de dados. Dessa forma, para a melhoria dos resultados obtidos na abordagem de classificação multiclasse, é fundamental que as redes sejam treinadas com um conjunto de dados muito maior, permitindo que realizem melhores abstrações dos atributos que diferem as várias possíveis classes.

5.2 CLASSIFICAÇÃO COM HIERARQUIA ENTRE AS TAREFAS – ELEMENTOS TÉCNICOS DIFERENTES

Nesta abordagem, a consideração da hierarquia entre as tarefas gerou uma primeira etapa de classificação binária, utilizada para a identificação entre os dois diferentes elementos técnicos possíveis no conjunto: Triple Axel (3A) e Quad Toe (4T). Essa etapa foi seguida por mais duas classificações binárias, utilizadas para a avaliação da qualidade de execução dos dois possíveis elementos considerados. Dessa forma, as atividades de identificação do elemento apresentado e avaliação da qualidade de execução foram avaliadas de forma independente, com treinamentos específicos para cada uma das atividades.

Nesta seção são apresentados os resultados obtidos para a atividade de identificação do elemento apresentado e, nas próximas duas seções, a atividade de avaliação da qualidade de execução para cada um dos dois tipos de elementos do conjunto. A tabela 3 apresenta os

valores de acurácia obtidos e as células marcadas em amarelo representam os resultados máximos alcançados.

Tabela 3 – Resultados obtidos com a abordagem de classificação binária para identificação do elemento apresentado

	Classificação binária (Triple Axel / Quad Toe)			
	ResNet-18		Xception	
	Atributos de alto nível	Atributos de baixo nível	Atributos de alto nível	Atributos de baixo nível
Fine Tree	46,7%	43,3%	46,7%	**80,0%**
Medium Tree	46,7%	43,3%	46,7%	80,0%
Coarse Tree	46,7%	43,3%	46,7%	80,0%
Linear Discriminant	-	63,3%	-	53,3%
Logistic Regression	56,7%	**76,7%**	46,7%	56,7%
Gaussian Naive Bayes	-	63,3%	-	50,0%
Kernel Naive Bayes	53,3%	53,3%	53,3%	53,3%
Linear SVM	56,7%	60,0%	46,7%	56,7%
Quadratic SVM	46,7%	53,3%	40,0%	43,3%
Cubic SVM	46,7%	56,7%	43,3%	43,3%
Fine Gaussian SVM	53,3%	53,3%	53,3%	53,3%
Medium Gaussian SVM	50,0%	53,3%	53,3%	36,7%
Coarse Gaussian SVM	53,3%	53,3%	53,3%	53,3%
Fine KNN	36,7%	50,0%	33,3%	43,3%
Medium KNN	40,0%	63,3%	40,0%	50,0%
Coarse KNN	53,3%	53,3%	53,3%	53,3%
Cosine KNN	43,3%	63,3%	40,0%	43,3%
Cubic KNN	43,3%	66,7%	40,0%	53,3%
Weighted KNN	33,3%	56,7%	40,0%	50,0%
Ensemble - Boosted Trees	53,3%	53,3%	53,3%	53,3%
Ensemble - Bagged Trees	53,3%	60,0%	26,7%	46,7%
Ensemble - Subspace Discriminant	43,3%	56,7%	56,7%	50,0%
Ensemble - Subspace KNN	23,3%	46,7%	33,3%	53,3%
Ensemble - RUSBoosted Trees	**70,0%**	63,3%	**66,7%**	60,0%

Fonte: Autora.
Legenda: Valores de acurácia apresentados por cada uma das técnicas de aprendizado de máquina avaliadas.

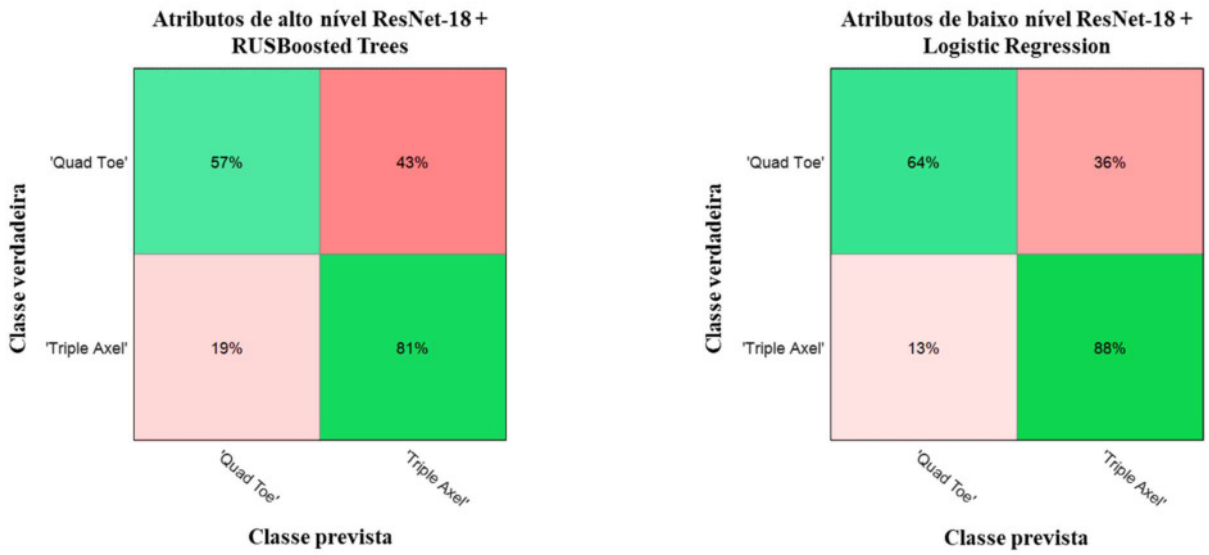
A partir da avaliação dos resultados apresentados na tabela 3, é possível observar que, na abordagem de classificação binária para identificação do elemento apresentado, o resultado geral foi bem melhor do que na abordagem com classificação multiclasse. Esse resultado demonstra que a consideração da hierarquia entre as tarefas e a simplificação do número de classes são boas abordagens para este problema, no qual o conjunto de dados é muito pequeno.

O melhor resultado geral foi obtido pelas técnicas de árvores de decisão com o uso dos atributos de baixo nível extraídos pela rede pré-treinada Xception, que apresentaram acurácia de 80%.

Nesta abordagem os resultados obtidos pelas duas redes utilizadas para extração de atributos também foram semelhantes, assim como ocorrido na abordagem de classificação multiclasse, apresentada na seção 4.1. Ao realizarmos a comparação entre os atributos de baixo e alto nível extraídos por uma mesma rede, é possível observar um pouco melhor nesta abordagem a vantagem do uso dos atributos de baixo nível para o treinamento, novamente confirmando a hipótese desses atributos serem mais adequados neste problema por descreverem a imagem com características mais primárias e menos contextuais, conforme abordado anteriormente.

As figuras 25 e 26 apresentam as matrizes de confusão das técnicas que apresentaram os melhores resultados com os atributos de baixo e alto nível extraídos pela rede ResNet e pela Xception, respectivamente.

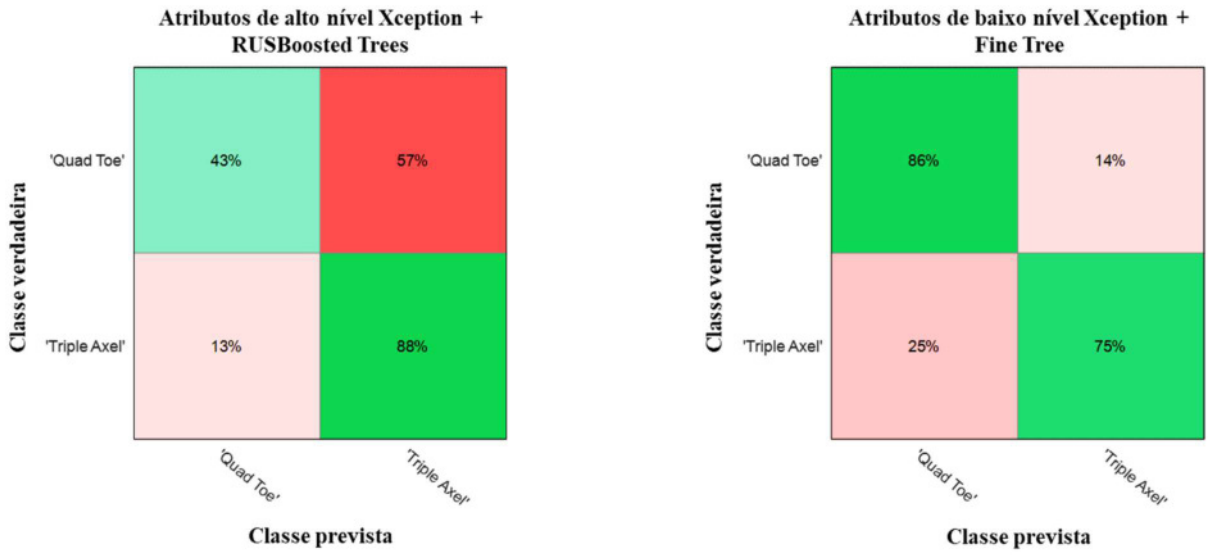
Figura 25 – Matrizes de confusão na abordagem de classificação binária para identificação do elemento apresentado, com atributos extraídos pela ResNet



Fonte: Autora.

Legenda: Matrizes de confusão das técnicas de aprendizado de máquina que apresentaram melhores resultados de acurácia nesta abordagem, com a extração de atributos de alto e baixo nível pela ResNet-18.

Figura 26 – Matrizes de confusão na abordagem de classificação binária para identificação do elemento apresentado, com atributos extraídos pela Xception



Fonte: Autora.
Legenda: Matrizes de confusão das técnicas de aprendizado de máquina que apresentaram melhores resultados de acurácia nesta abordagem, com a extração de atributos de alto e baixo nível pela Xception.

Através da avaliação da figura 25 é possível observar que os resultados obtidos na identificação do elemento Triple Axel foram melhores do que na identificação do elemento Quad Toe. Esse comportamento pode estar relacionado com uma maior complexidade dos atributos apresentados nas imagens do elemento Quad Toe e ao fato do conjunto de dados possuir uma quantidade ligeiramente maior de imagens do elemento Triple Axel (48 imagens) do que do elemento Quad Toe (42 imagens). O uso de um conjunto de dados maior e mais equilibrado seria uma boa abordagem para melhorar o treinamento dos classificadores, de forma que eles conseguissem abstrair melhores e mais complexas características através dos atributos observados.

Observando os resultados obtidos com os atributos extraídos pela rede Xception (figura 26), o comportamento acima descrito só foi observado no caso do treinamento através dos atributos de alto nível. Para o treinamento através dos atributos de baixo nível o resultado foi mais equilibrado entre a identificação dos dois elementos, apresentando acurácia até superior para a identificação do Quad Toe. As três técnicas de árvores de decisão avaliadas (fine tree, medium tree e coarse tree) apresentaram resultados extremamente consistentes, com mesmos valores finais de acurácia e distribuições nas matrizes de confusão.

Para uma avaliação completa do uso da classificação binária neste problema, é necessário levar em consideração tanto os resultados obtidos na identificação dos elementos,

já apresentados nesta seção, quanto os obtidos na avaliação da qualidade de execução dos elementos, que serão detalhados e discutidos na sequência.

5.3 CLASSIFICAÇÃO COM HIERARQUIA ENTRE AS TAREFAS – QUALIDADE DE EXECUÇÃO DO TRIPLE AXEL

A avaliação da qualidade de execução do Triple Axel é tratada nesta abordagem como um problema de classificação binária, no qual os classificadores são treinados em um subconjunto de imagens, composto apenas pelas imagens do elemento Triple Axel existentes no conjunto completo de dados. Assim, o subconjunto utilizado possui 48 imagens no total, sendo 27 de execuções com pontuação acima (3A pos) e 21 de execuções com pontuação abaixo (3A neg) do valor base do elemento. A tabela 4 apresenta os valores de acurácia obtidos e as células marcadas em amarelo representam os resultados máximos alcançados.

Tabela 4 – Resultados obtidos com a abordagem de classificação binária para avaliação da qualidade de execução do Triple Axel

	Classificação binária (Triple Axel+ / Triple Axel-)			
	ResNet-18		Xception	
	Atributos de alto nível	Atributos de baixo nível	Atributos de alto nível	Atributos de baixo nível
Fine Tree	50,0%	56,3%	6,3%	68,8%
Medium Tree	50,0%	56,3%	6,3%	68,8%
Coarse Tree	50,0%	56,3%	6,3%	68,8%
Linear Discriminant	-	37,5%	-	62,5%
Logistic Regression	43,8%	31,3%	**68,8%**	37,5%
Gaussian Naive Bayes	-	50,0%	-	43,8%
Kernel Naive Bayes	56,3%	50,0%	56,3%	56,3%
Linear SVM	62,5%	25,0%	37,5%	56,3%
Quadratic SVM	**81,3%**	37,5%	43,8%	50,0%
Cubic SVM	56,3%	25,0%	56,3%	50,0%
Fine Gaussian SVM	56,3%	56,3%	56,3%	56,3%
Medium Gaussian SVM	56,3%	37,5%	56,3%	50,0%
Coarse Gaussian SVM	56,3%	56,3%	56,3%	56,3%
Fine KNN	75,0%	25,0%	62,5%	50,0%
Medium KNN	37,5%	50,0%	31,3%	43,8%
Coarse KNN	56,3%	56,3%	56,3%	56,3%
Cosine KNN	31,3%	31,3%	31,3%	37,5%
Cubic KNN	50,0%	50,0%	31,3%	37,5%
Weighted KNN	56,3%	31,3%	37,5%	56,3%
Ensemble - Boosted Trees	56,3%	56,3%	56,3%	56,3%
Ensemble - Bagged Trees	43,8%	12,5%	37,5%	**75,0%**
Ensemble - Subspace Discriminant	68,8%	31,3%	50,0%	37,5%
Ensemble - Subspace KNN	56,3%	43,8%	56,3%	50,0%
Ensemble - RUSBoosted Trees	62,5%	**68,8%**	43,8%	68,8%

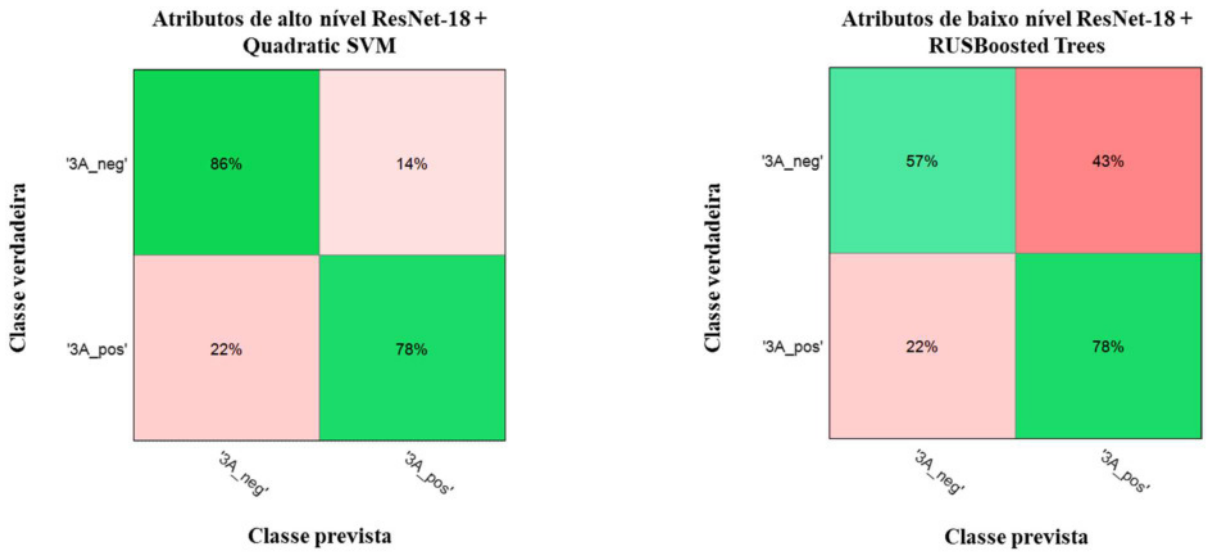
Fonte: Autora.
Legenda: Valores de acurácia apresentados por cada uma das técnicas de aprendizado de máquina avaliadas.

Avaliando a tabela 4 é possível observar que os resultados obtidos nesta abordagem continuam sendo melhores do que os obtidos na abordagem de classificação multiclasse. O melhor resultado geral foi obtido pela técnica *Quadratic SVM* com o uso dos atributos de alto nível extraídos pela rede pré-treinada ResNet, que apresentou acurácia de 81,3%. Nesta abordagem podemos observar pela primeira vez o melhor valor de acurácia ter sido atingido através do treinamento com atributos de alto nível extraídos pela rede ResNet, ao invés do treinamento com atributos de baixo nível extraídos pela rede Xception.

Para os treinamentos com os atributos de baixo nível observa-se que, tanto na extração pela ResNet quanto na extração pela Xception, as técnicas que apresentaram melhor acurácia foram técnicas que utilizam conjuntos de árvores de decisão (*ensemble bagged trees* e *ensemble RUSboosted trees*), reforçando o bom desempenho apresentado por esses classificadores nas abordagens apresentadas anteriormente.

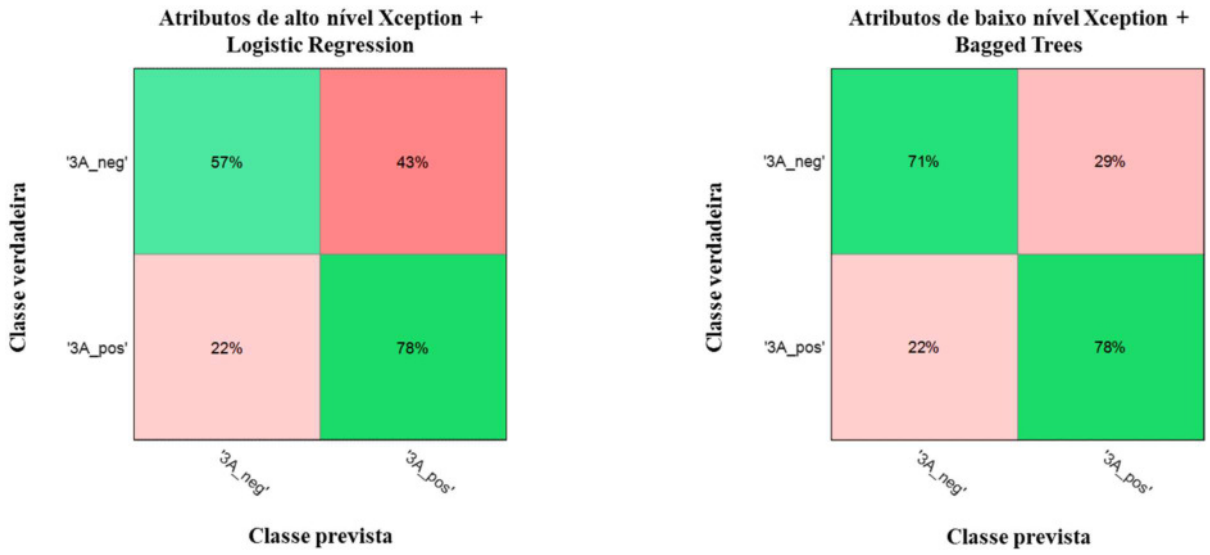
As figuras 27 e 28 apresentam as matrizes de confusão das técnicas que apresentaram os melhores resultados com os atributos de baixo e alto nível extraídos pela rede ResNet e pela Xception, respectivamente.

Figura 27 – Matrizes de confusão na abordagem de classificação binária para avaliação da qualidade de execução do Triple Axel, com atributos extraídos pela ResNet



Fonte: Autora.
Legenda: Matrizes de confusão das técnicas de aprendizado de máquina que apresentaram melhores resultados de acurácia nesta abordagem, com a extração de atributos de alto e baixo nível pela ResNet-18.

Figura 28 – Matrizes de confusão na abordagem de classificação binária para avaliação da qualidade de execução do Triple Axel, com atributos extraídos pela Xception



Fonte: Autora.
Legenda: Matrizes de confusão das técnicas de aprendizado de máquina que apresentaram melhores resultados de acurácia nesta abordagem, com a extração de atributos de alto e baixo nível pela Xception.

Através da avaliação da figura 27, na qual os atributos foram extraídos pela rede ResNet, é possível observar que os resultados obtidos com os atributos de alto nível foram mais equilibrados entre as classes do que os resultados obtidos com os atributos de baixo nível, no qual a matriz de confusão apresenta uma dificuldade do classificador com as imagens do Triple Axel que tiveram pontuação abaixo do valor base do elemento. Observando a figura 28 nota-se que o classificador também encontrou dificuldades com as imagens da categoria Triple Axel-, no entanto este comportamento ocorreu com o uso dos atributos de alto nível das imagens, ao invés dos atributos de baixo nível.

Como o conjunto de dados utilizado é bastante pequeno, é difícil determinar se os resultados obtidos nessa abordagem com a rede ResNet e os atributos de alto nível apresentam uma mudança na tendência de comportamento quando avaliada a qualidade da ação, ou se é um fato isolado que possa ter sido favorecido por algum atributo presente nas imagens do Triple Axel, que seja ressaltado nas camadas mais profundas da ResNet.

Na seção seguinte serão apresentados os resultados da avaliação da qualidade de execução do elemento Quad Toe.

5.4 CLASSIFICAÇÃO COM HIERARQUIA ENTRE AS TAREFAS – QUALIDADE DE EXECUÇÃO DO QUAD TOE

Assim como apresentado na seção anterior para o elemento Triple Axel, a avaliação da qualidade de execução do Quad Toe é tratada nesta abordagem como um problema de classificação binária. O subconjunto de imagens utilizado para o treinamento dos classificadores possui apenas as imagens do elemento Quad Toe existentes no conjunto completo de dados, sendo composto por 42 imagens no total, das quais 27 apresentam execuções com pontuação acima (4T pos) e 15 apresentam execuções com pontuação abaixo (4T neg) do valor base do elemento. A tabela 5 apresenta os valores de acurácia obtidos e as células marcadas em amarelo representam os resultados máximos alcançados.

Tabela 5 – Resultados obtidos com a abordagem de classificação binária para avaliação da qualidade de execução do Quad Toe

	Classificação binária (Quad Toe+ / Quad Toe-)			
	ResNet-18		Xception	
	Atributos de alto nível	Atributos de baixo nível	Atributos de alto nível	Atributos de baixo nível
Fine Tree	71,4%	28,6%	35,7%	71,4%
Medium Tree	71,4%	28,6%	35,7%	71,4%
Coarse Tree	71,4%	28,6%	35,7%	71,4%
Linear Discriminant	-	50,0%	-	28,6%
Logistic Regression	78,6%	57,1%	50,0%	57,1%
Gaussian Naive Bayes	-	57,1%	-	**85,7%**
Kernel Naive Bayes	64,3%	57,1%	**64,3%**	64,3%
Linear SVM	42,9%	57,1%	28,6%	42,9%
Quadratic SVM	50,0%	57,1%	50,0%	64,3%
Cubic SVM	57,1%	57,1%	35,7%	42,9%
Fine Gaussian SVM	64,3%	64,3%	**64,3%**	64,3%
Medium Gaussian SVM	64,3%	64,3%	**64,3%**	64,3%
Coarse Gaussian SVM	64,3%	64,3%	**64,3%**	64,3%
Fine KNN	28,6%	57,1%	42,9%	35,7%
Medium KNN	57,1%	64,3%	50,0%	57,1%
Coarse KNN	64,3%	64,3%	**64,3%**	64,3%
Cosine KNN	50,0%	35,7%	50,0%	50,0%
Cubic KNN	57,1%	64,3%	50,0%	57,1%
Weighted KNN	57,1%	64,3%	57,1%	64,3%
Ensemble - Boosted Trees	64,3%	64,3%	**64,3%**	64,3%
Ensemble - Bagged Trees	**85,7%**	35,7%	42,9%	50,0%
Ensemble - Subspace Discriminant	42,9%	**71,4%**	50,0%	42,9%
Ensemble - Subspace KNN	64,3%	35,7%	42,9%	57,1%
Ensemble - RUSBoosted Trees	50,0%	14,3%	50,0%	78,6%

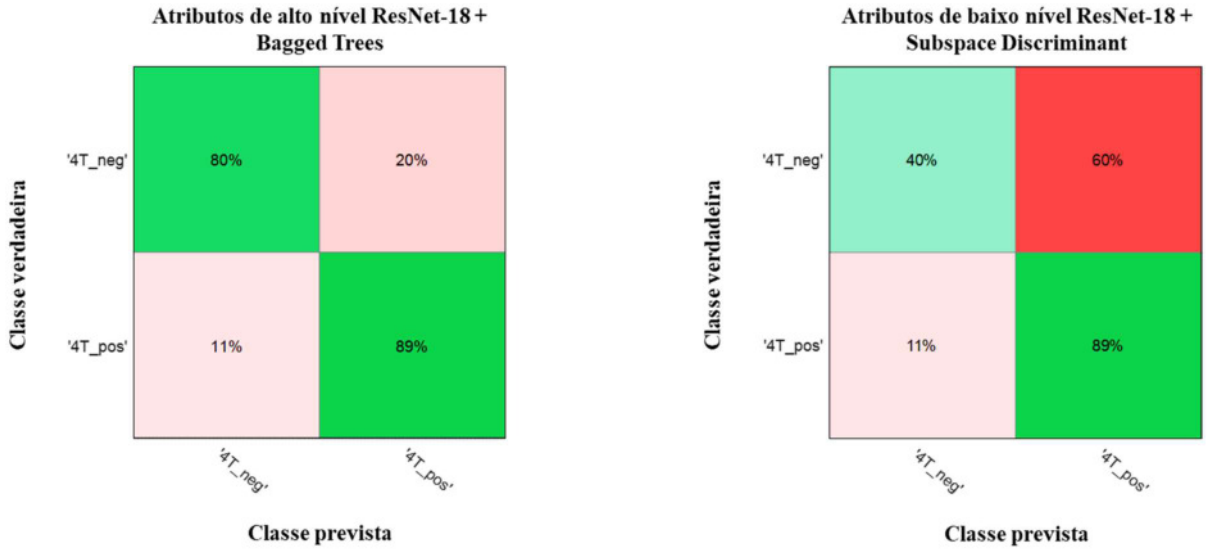
Fonte: Autora.
Legenda: Valores de acurácia apresentados por cada uma das técnicas de aprendizado de máquina avaliadas.

Os resultados apresentados na tabela 5 confirmam que a abordagem através de classificação binária e considerando a hierarquia entre as tarefas alcança melhores níveis de acurácia para o problema avaliado. O melhor resultado geral foi de 85,7% de acurácia, obtido tanto pela técnica *Ensemble – Bagged Trees* com o uso dos atributos de alto nível extraídos pela rede pré-treinada ResNet, quanto pela técnica *Gaussian Naïve Bayes* com o uso dos atributos de baixo nível extraídos pela rede pré-treinada Xception.

Assim como observado na avaliação da qualidade de execução do Triple Axel, nessa abordagem os resultados obtidos com a extração de atributos pela rede ResNet também foram melhores com o uso dos atributos de alto nível, ao invés dos de baixo nível, reforçando a hipótese de uma possível mudança na tendência de comportamento quando avaliada a qualidade da ação. No caso dos resultados obtidos com a extração de atributos pela Xception, não foi observada mudança de comportamento entre as diferentes abordagens avaliadas, sendo os melhores resultados obtidos com o uso dos atributos de baixo nível, tanto para identificação do elemento apresentado quanto para a avaliação da qualidade de execução.

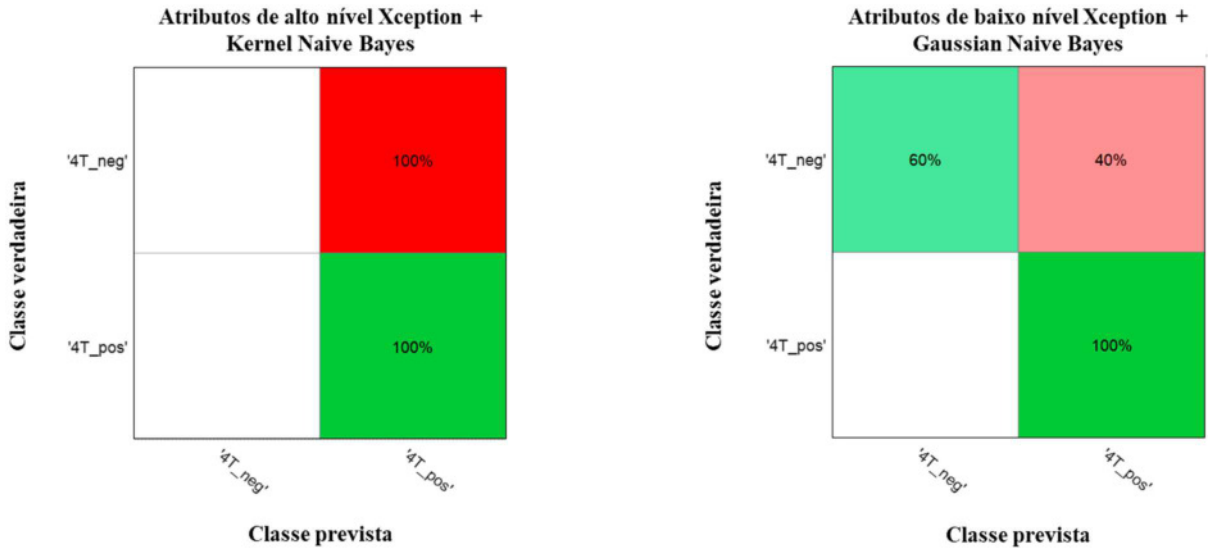
As figuras 29 e 30 apresentam as matrizes de confusão das técnicas que apresentaram os melhores resultados com os atributos de baixo e alto nível extraídos pela rede ResNet- e pela Xception, respectivamente.

Figura 29 – Matrizes de confusão na abordagem de classificação binária para avaliação da qualidade de execução do Quad Toe, com atributos extraídos pela ResNet



Fonte: Autora.
Legenda: Matrizes de confusão das técnicas de aprendizado de máquina que apresentaram melhores resultados de acurácia nesta abordagem, com a extração de atributos de alto e baixo nível pela ResNet-18.

Figura 30 – Matrizes de confusão na abordagem de classificação binária para avaliação da qualidade de execução do Quad Toe, com atributos extraídos pela Xception



Fonte: Autora.
Legenda: Matrizes de confusão das técnicas de aprendizado de máquina que apresentaram melhores resultados de acurácia nesta abordagem, com a extração de atributos de alto e baixo nível pela Xception.

Avaliando as figuras 29 e 30 é possível observar que a maior dificuldade dos classificadores foi com as imagens do Quad Toe que tiveram pontuação abaixo do valor base do elemento (4T neg). Esse mesmo comportamento foi apresentado tanto com o uso dos atributos de alto nível, quanto com o uso dos atributos de baixo nível, extraídos pela ResNet-18 e pela Xception. Uma possível causa para esse comportamento é o desbalanceamento no subconjunto de dados utilizado para o treinamento da rede, que possuía 15 imagens de execuções com pontuação abaixo do valor base, frente a 27 imagens de execuções com pontuação acima do valor base.

Uma discussão geral sobre os resultados obtidos nas diferentes abordagens avaliadas, bem como possíveis soluções a serem abordadas no futuro, serão apresentadas na próxima seção.

5.5 DISCUSSÃO GERAL SOBRE OS RESULTADOS

A partir de uma análise sobre os resultados gerais obtidos em todas as abordagens avaliadas e apresentadas nas seções anteriores, algumas observações são possíveis.

A proposta de realizar atividades de classificação através de imagens *stop-motion* (ou quadra a quadro), tanto para a identificação quanto para a avaliação da qualidade de execução

dos elementos, mostrou-se promissora para esta tarefa. Além de reduzir a complexidade do problema em termos de arquitetura de rede, aprendizagem e complexidade computacional, essas imagens já são amplamente aceitas no domínio dos esportes, permitindo ao público um melhor entendimento dos elementos e aos atletas e técnicos um recurso valioso para o aprimoramento das performances. Adicionalmente, já existem no mercado softwares capazes de criar as imagens quadro a quadro automaticamente, apenas com a indicação de alguns parâmetros pelo usuário. O uso desses softwares é uma possibilidade de otimização da proposta apresentada nesta dissertação, já que os esforços com a montagem manual das imagens seriam reduzidos fortemente.

Um ponto crítico e desafiador deste trabalho foi o reduzido tamanho do conjunto de dados, devido à especificidade da tarefa proposta e da complexidade anteriormente citada para a montagem manual de todas as imagens. O treinamento de redes profundas demanda milhares de imagens rotuladas, o que inviabilizou a utilização desta técnica para as atividades de classificação. No entanto, a abordagem proposta de realizar uma combinação da técnica de aprendizagem profunda para a extração de atributos, com o uso de outras técnicas de aprendizado de máquina para as atividades de classificação, se mostrou efetiva. Essa abordagem permitiu a união do ponto forte dessas técnicas e uma solução ao problema apresentado, através da integração do rico mapeamento de atributos realizado pelas redes profundas, com a demanda de conjuntos de dados menores pelas demais técnicas classificadoras.

Ao comparar a abordagem do problema por meio da classificação multiclasse com a abordagem através de diversas classificações binárias é possível observar que, conforme esperado, as abordagens binárias apresentaram resultados melhores, com acurácias entre 65% e 85%, frente às acurácias alcançadas com a classificação multiclasse, entre 30% e 45%. Certamente esse resultado também foi influenciado pelo tamanho do conjunto, de forma que conjuntos de dados maiores poderiam fornecer aos classificadores mais e melhores informações de diferenciação entre as classes possíveis, beneficiando ambos os tipos de classificação, principalmente a multiclasse.

Em termos de extração de atributos, ambas as redes avaliadas (ResNet e Xception) levaram a resultados muito semelhantes, com pequena vantagem para a ResNet. Essa observação demonstra que redes com diferentes arquiteturas são capazes de fornecer uma representação efetiva dos dados, de forma que a abordagem proposta neste trabalho não se limita a um único tipo de arquitetura de aprendizagem de atributos. No entanto, é importante realçar que dificilmente valores de acurácia na escala de 85% teriam sido alcançados com

uma escolha manual de atributos das imagens, especialmente em um conjunto de dados tão pequeno. Isso comprova a imensa capacidade de aprendizagem de atributos das redes neurais profundas, o que resulta nos ótimos resultados de classificação que vêm sendo alcançados por essas redes atualmente.

Em relação aos diferentes níveis de atributos extraídos e utilizados nas tarefas de classificação, os atributos de baixo nível se mostraram mais adequados para essa tarefa. Conforme já comentado anteriormente, esse comportamento se deve principalmente ao fato de que as redes ResNet e Xception foram pré-treinadas no conjunto de dados ImageNet, que não é um conjunto composto por imagens de patinação artística, nem focado em esportes em geral. O conjunto possui grande quantidade de imagens de diversos temas, principalmente imagens naturais e de animais, o que torna os atributos de alto nível menos adequados para este problema, já que estes atributos descrevem a imagem em um nível de conhecimento e contexto. Um ponto interessante observado foi o comportamento inverso apresentado pela ResNet nas atividades de avaliação da qualidade de execução dos elementos, nas quais os melhores resultados foram obtidos com os atributos de alto nível. No entanto, a realização de mais testes com conjuntos de dados maiores seria necessária para confirmar e avaliar melhor essa possível tendência.

Analisando os resultados obtidos através dos diferentes algoritmos classificadores avaliados é possível observar que os algoritmos baseados em árvores de decisão, tanto os simples quanto os *ensembles*, tiveram resultados gerais muito superiores aos demais algoritmos avaliados. Normalmente algoritmos baseados em lógica, como as árvores de decisão, apresentam resultados melhores quando lidam com atributos discretos/categóricos, o que pode indicar uma característica predominante dos atributos extraídos pelas redes profundas no conjunto de dados deste problema. Além disso, destaca-se o curto tempo computacional necessário para o treinamento das técnicas de árvores de decisão, sendo este um ponto positivo adicional aos valores de acurácia alcançados.

Os classificadores naïve Bayes, SVM, kNN, regressão logística e análise discriminante linear apresentaram bons resultados apenas em algumas avaliações isoladas, não sendo os mais indicados no geral para o problema proposto. Algoritmos kNN costumam ser bastante sensíveis à atributos irrelevantes, o que pode facilmente justificar os resultados apresentados, já que foram usados centenas de atributos extraídos pelas redes profundas e, certamente, muitos deles não seriam os mais adequados para estes algoritmos. No caso das máquinas de vetores de suporte (SVM), normalmente são necessários conjuntos de dados grandes para que

o algoritmo atinja sua precisão máxima de previsão, o que não estava disponível nesse problema e pode ter reduzido a precisão do algoritmo.

Por fim, é importante citar que através da avaliação das matrizes de confusão é possível observar que o desbalanceamento do conjunto de dados, especialmente no caso das imagens do elemento Quad Toe, interferiu no resultado. Um conjunto de dados mais balanceado, especialmente no caso de um conjunto de dados pequeno como o deste trabalho, certamente seria um fator de melhoria na performance dos classificadores.

As várias avaliações realizadas na proposta deste trabalho confirmam o conhecimento atual de que não há um único classificador que seja melhor que todos os outros, em todos os conjuntos de dados. Cada um dos classificadores possui características específicas, que devem ser analisadas e testadas em cada problema, de forma a se identificar quais apresentam maior precisão para o problema abordado.

6 CONCLUSÃO

Este trabalho apresentou uma proposta para a avaliação de movimentos na patinação artística, através da identificação e interpretação de resultados automáticas. O domínio da patinação artística é bastante desafiador, pois une características técnicas e de avaliação objetiva com características artísticas e de avaliação mais subjetiva. O sistema de avaliação é complexo, com grande quantidade de critérios para garantir uma avaliação justa e transparente das apresentações. Além disso, o domínio reúne uma série de desafios na área de visão computacional, como planos de fundo dinâmicos, câmeras em plataformas móveis e áreas de oclusões, o que torna as atividades de identificação e avaliação da execução mais complexas. Diante dos diversos desafios que envolvem esse domínio, uma abordagem simplificada do problema foi adotada, de forma a avaliar algumas propostas iniciais e identificar possíveis trabalhos futuros.

A abordagem adotada envolveu três etapas de implementação principais: a elaboração do conjunto de dados, a extração de atributos das imagens do conjunto e a classificação das imagens. Para isso, foram utilizadas técnicas de segmentação de imagens, aprendizagem profunda e técnicas de classificação para aprendizado supervisionado de máquina.

O conjunto de dados elaborado é composto por imagens que apresentam execuções em *stop-motion* (quadro a quadro) de dois elementos técnicos selecionados para avaliação neste trabalho, o Triple Axel e o Quad Toe. As imagens em *stop-motion* foram escolhidas pois conseguem representar em uma única imagem a sequência completa de execução de um elemento, diminuindo a quantidade de dados de entrada para as tarefas de treinamento e a complexidade.

O uso de algoritmos classificadores para treinamento supervisionado demanda que os dados de entrada sejam descritos através de seus atributos. Dessa forma, as imagens que compõem o conjunto foram utilizadas como dados de entrada de redes de aprendizagem profunda pré-treinadas, que possuem grande capacidade de seleção e extração de atributos. Os atributos extraídos pelas redes profundas, por sua vez, foram combinados com os rótulos das imagens e utilizados como dados de entrada para treinamento de algoritmos classificadores.

Para o treinamento dos classificadores, diversas abordagens diferentes foram adotadas, de forma a permitir uma vasta avaliação do problema e das melhores soluções encontradas. A extração de atributos se deu através de duas redes profundas de arquiteturas diferentes, além de serem extraídos atributos em dois níveis diferentes de cada rede, de forma a serem obtidos atributos de baixo e de alto nível das imagens do conjunto de dados. Para as atividades de

classificação, diferentes técnicas foram avaliadas, como árvores de decisão, análise discriminante linear, naïve Bayes, k vizinhos mais próximos, máquinas de vetores de suporte, conjuntos de classificadores (*ensembles*) e regressão logística. Além disso, as tarefas de classificação foram abordadas de duas maneiras diferentes em relação a hierarquia entre elas, gerando dois diferentes cenários de classificação, um deles com uma tarefa de classificação multiclasse e o outro com três tarefas de classificações binárias.

Todos os modelos treinados em suas diversas abordagens foram avaliados através do método de validação cruzada *leave-one-out*, de forma a verificar com maior exatidão o nível de precisão alcançado por cada modelo. Os resultados foram expressos em termo da acurácia dos modelos e foram avaliadas as matrizes de confusão das técnicas que apresentaram melhores resultados, permitindo uma análise detalhada do comportamento desses classificadores quando treinados com o conjunto de dados apresentado.

Os resultados obtidos para as diversas abordagens permitiram uma avaliação rica do problema e proposta de solução. A abordagem com tarefa de classificação multiclasse foi a que apresentou resultados inferiores, principalmente devido à pequena quantidade de dados disponíveis no conjunto para o treinamento dos diversos classificadores avaliados. Dessa forma, foi possível concluir que as abordagens de classificação binária alcançam melhores resultados com conjuntos de dados pequenos. Dentre os diversos algoritmos classificadores avaliados, os que utilizam como base as árvores de decisão apresentaram, no geral, desempenho superior neste problema, o que pode estar relacionado ao tipo das imagens do conjunto e aos atributos que foram extraídos pelas redes profundas. Ambas as redes profundas avaliadas, ResNet e Xception, apresentaram resultados excelentes na identificação e extração de atributos, confirmando o poder das redes de aprendizagem profunda para esta tarefa. Entre os níveis de atributos extraídos e utilizados para o treinamento dos classificadores, os atributos de baixo nível apresentaram resultados melhores de forma geral, o que está bastante associado ao fato de as redes profundas que realizaram a extração terem sido pré-treinadas em um conjunto de dados bastante diferente do proposto neste trabalho.

Diante de todas as análises realizadas com os resultados obtidos, é possível considerar que a proposta apresentada conseguiu encontrar resultados satisfatórios e promissores em várias das abordagens realizadas. Foram também observados pontos de melhoria que geram algumas novas propostas para trabalhos futuros, de forma a aprimorar os resultados atualmente obtidos. A criação de um conjunto de dados significativamente maior é de extrema importância para a melhoria dos resultados obtidos com as atividades de aprendizado de máquina. Diante de um conjunto maior de dados, seria interessante investigar também o uso

das redes neurais profundas não apenas como extratores de atributos, mas também como modelos classificadores, o que não foi possível na proposta inicial devido à pequena quantidade disponível de dados. Adicionalmente, uma proposta que pode trazer melhoria dos resultados é a realização de um pré-treinamento das redes profundas através de um conjunto de dados composto por imagens de esportes.

Em propostas futuras que pretendam abordar o problema de forma mais completa, seria interessante que os modelos fossem capazes de avaliar os elementos executados durante a apresentação tanto de forma individual, quanto como um conjunto (incluindo todos os elementos executados, a música e a pista), para que sejam observados também os critérios que relacionam os elementos entre si e avaliam os aspectos abstratos da apresentação.

REFERÊNCIAS

(1) AGGARWAL, J. K.; RYOO, M. S. Human activity analysis: a review. **Journal ACM Computing Surveys (CSUR)**. ACM New York, v. 43, n. 3, article 16, 2011.

(2) ALY, M. Survey on multiclass classification methods. **Technical Report**. 2005.

(3) AMAZON WEB SERVICES. **Amazon Machine Learning: Guia do desenvolvedor.** 2016. Disponível em <https://docs.aws.amazon.com/machine-learning/latest/dg/ machinelearning-dg.pdf>. Acesso em 03 ago. 2019.

(4) BASHEER, I. A.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design and application. **Journal of Microbiological Methods**. Elsevier, v. 43, n. 1, 3-31, 2000.

(5) BAYES, T. An essay towards solving a problem in the doctrine of chances. **Philosophical Transactions of the Royal Society.** The Royal Society Publishing, v. 53, 370-418, 1763.

(6) BISHOP, C. M. **Pattern Recognition and Machine Learning.** 1 ed. New York, USA: Springer-Verlag, 2006.

(7) BOBICK, A. F. Movement, activity and action: the role of knowledge in the perception of motion. **Journal Philosophical transactions of the Royal Society of London. Series B, Biological sciences**. The Royal Society Publishing, v. 352, n. 1358, 1257-1265, 1997.

(8) BOYKOV, Y. Y.; JOLLY, M.-P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), 8., 2001, Vancouver. **Proceedings**… Vancouver, CAN, 2001. p. 105-112.

(9) BURGES, C. J. C. A Tutorial on Support Vector Machines for Pattern Recognition. **Data Mining and Knowledge Discovery**. USA: Kluwer Academics Publishers, v. 2, n. 2, 121-167, Jun. 1998.

(10) CHOLLET, F. Xception: deep learning with depthwise separable convolutions. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2017, Honolulu. **Proceedings**… Honolulu, USA, 2017, p. 1800-1807.

(11) CONFEDERAÇÃO BRASILEIRA DE DESPORTOS NO GELO. **Patinação artística**. 2019. Disponível em <http://www.cbdg.org.br/modalidades/patinacao-artistica/>. Acesso em: 20 set. 2019.

(12) COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**. USA: IEEE, v. 13, n. 1, 21-27, Jan. 1967.

(13) CRISTIANINI, N.; SHAWE-TAYLOR, J. **An introduction to Support Vector Machines and other kernel-based learning methods**. 1. ed. New York, USA: Cambridge University Press, 2000.

(14) DATA SCIENCE DOJO. **One Versus One vs. One Versus All in Classification Models.** 2019. Disponível em <https://tutorials.datasciencedojo.com/classification-one-vs-one-vs-all/>. Acesso em 01 out. 2019.

(15) DENG, J. ET AL. ImageNet: A large-scale hierarchical image database. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2009, Miami. **Proceedings**… Miami, USA, 2009, p. 248-255.

(16) EDGE CRUNCH. **What is figure skating?** 2017. Disponível em <https://medium.com/edge-crunch/what-is-figure-skating-ac0f456fd4a4> Acesso em 30 set. 2019.

(17) FORSYTH, D. A. ET AL. Computational studies of human motion: part 1, tracking and motion synthesis. **Journal Foundations and Trends in Computer Graphics and Vision**. Now Publishers Inc., v. 1, n. 2-3, 77-254, 2005.

(18) FRIEDMAN, J. H. Regularized Discriminant Analysis. **Journal of the American Statistical Association.** Taylor & Francis, v. 84, n. 405, 165-175, 1989.

(19) GOLDBERG, D. E. **Genetic algorithms in search optimization and machine learning**. 1. ed. Boston, USA: Addison-Wesley Longman Publishing Co., 1989.

(20) GONZALEZ, R. C.; WOODS, R. E. **Digital Imaging Processing**. 2. ed. New York, USA: Prentice Hall, 2002.

(21) HE, K. ET AL. Deep residual learning for image recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2016, Las Vegas. **Proceedings**… Las Vegas, USA, 2016, p. 770-778.

(22) HECKERMAN, D. A tutorial on learning with Bayesian networks. In: HOLMES, D. E., JAIN, L. C. **Innovations in Bayesian Networks**. **Studies in Computational Intelligence.** Berlin, DEU: Springer, 2008, v. 156. p. 33-82.

(23) INTERNATIONAL SKATING UNION. **Results**. Disponível em <https://www.isu.org/figure-skating/entries-results/fsk-results>. Acesso em 15 jan. 2019.

(24) INTERNATIONAL SKATING UNION. **Special regulations & technical rules – Single & pair skating and ice dance**. 2018. Disponível em <https://www.isu.org/figure-skating>. Acesso em 30 jun. 2019.

(25) INTERNATIONAL SKATING UNION. **ISU Figure skating media guide 2018/19**. 2019. Disponível em <https://www.isu.org/figure-skating>. Acesso em 25 jul. 2019.

(26) IVANOV, Y.; BOBICK, A. F. Recognition of visual activities and interactions by stochastic parsing. **IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)**. IEEE, v. 22, n. 8, p. 852-872, Ago. 2000.

(27) JAIN, A. K.; MAO, J.; MOHIUDDIN, K. Artificial neural networks: a tutorial. **IEEE Computer**. USA: IEEE Publisher, v. 29, n. 3, 31-44, Mar. 1996.

(28) JARDIM, D.; NUNES, L. M.; DIAS, M. S. Human activity recognition and prediction. In: DOCTORAL CONSORTIUM – DCPRAM, 2015, Lisbon. **Proceedings…** Lisbon, PRT, 2015. p. 24-32.

(29) KARPATHY, A. ET AL. Large-scale video classification with convolutional neural networks. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2014, Columbus. **Proceedings**… Columbus, USA, 2014. p. 1725-1732.

(30) KOHAVI, R. A. study of cross-validation and bootstrap for accuracy estimation and model selection. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE (IJCAI), 14., 1995, Montreal. **Proceedings**… Montreal, CAN, 1995, p. 1137-1145.

(31) KOTSIANTIS, S. B. Supervised Machine Learning: A Review of Classification Techniques. In: CONFERENCE ON EMERGING ARTIFICIAL INTELLIGENCE APPLICATIONS IN COMPUTER ENGINEERING, 2007, Amsterdam. **Proceedings**… Amsterdam, NLD: IOS Press, 2007. p. 3-24.

(32) LIN, T.-Y. ET AL. Microsoft COCO: Common Objects in Context. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 13., 2014, Zurich. **Proceedings**... Zurich, CHE, 2014. p. 740-755.

(33) MATHWORKS. **Introducing Machine Learning**. 2016. Disponível em < https://www.mathworks.com/campaigns/offers/machine-learning-with-matlab.html#> Acesso em 13 jun. 2019.

(34) MATHWORKS. **Introducing Deep Learning with Matlab**. 2018. Disponível em <https://www.mathworks.com/campaigns/offers/deep-learning-with-matlab.html>. Acesso em 13 jun. 2019.

(35) MCCULLOCH, W. S., PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics.** Kluwer Academic Publishers, v. 5, n. 4, 115-133, Dec.1943.

(36) MIKA, S. ET AL. Fisher discriminant analysis with kernel. In: NEURAL NETWORKS FOR SIGNAL PROCESSING IX: IEEE SIGNAL PROCESSING SOCIETY WORKSHOP, 2009, Madison. **Proceedings**… Madison, USA: IEEE, 2009. p. 41-48.

(37) MOESLUND, T. B.; HILTON, A.; KRÜGER, V. A survey of advances in vision-based human motion capture and analysis. **Journal Computer Vision and Image Understanding - Special issue on modeling people: vision-based understanding of a person's shape, appearance, movement, and behaviour**. Elsevier Science Inc., v. 104, n. 2, 90-126, 2006.

(38) OLYMPIC CHANNEL. **The jumps, spins and turns of figure skating**. 2018. Disponível em <https://www.olympicchannel.com/en/stories/news/detail/the-jumps-spins-and-turns-of-figure-skating/>. Acesso em 27 set. 2019.

(39) OLYMPIC GAMES. **Figure skating**. 2019. Disponível em <http://www.olympic.org/figure-skating>. Acesso em: 20 set. 2019.

(40) PARK, S.; AGGARWAL, J. K. A hierarchical Bayesian network for event recognition of human actions and interactions. **Multimedia Systems.** Berlin, DEU: Springer-Verlag, 2004, v. 10, n. 2, p. 164-179.

(41) PARMAR, P., MORRIS, B. T. Learning to score olympic events. In: COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (CVPRW), 2017, Honolulu. **IEEE Conference on**… Honolulu, USA, 2017. p. 76-84.

(42) PIRSIAVASH, H.; VONDRICK, C.; TORRALBA, A. Assessing the quality of actions. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 13., 2014, Zurich. **Proceedings**… Zurich, CHE, 2014. p. 556-571.

(43) POPPE, R. A. survey on vision-based human action recognition. **Journal Image and Vision Computing**. Elsevier, v. 28, n. 6, 976-990, 2010.

(44) POPPE, R. Vision-based human motion analysis: an overview. **Journal Computer Vision and Image Understanding**. Elsevier Science Inc., v. 108, n. 1-2, 4-18, 2007.

(45) RAMBABU, C.; CHAKRABARTI, I.; MAHANTA, A. Flooding-based watershed algorithm and its prototype hardware architecture. **IEE Proceedings – Vision, Image and Signal Processing**. IET, v. 151, n. 3, p. 224-234. Jul. 2004.

(46) RAUT, S. A. ET AL. Image Segmentation – A State-Of-Art Survey for Prediction. In: INTERNATIONAL CONFERENCE ON ADVANCED COMPUTER CONTROL (ICACC), 2009, Singapore. **Proceedings**… Singapore, SGP, 2009. p. 420-424.

(47) ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**. American Psychological Association, v. 65, n. 6. 386-408, 1958.

(48) SKATE SWEDEN. **Jumps – an introduction**. 2019. Disponível em <http://skatesweden.se/en/figureskatingatoz/jumps-an-introduction/>. Acesso em 01 out. 2019.

(49) STATISTICS SOLUTIONS. **What is Logistic Regression?** 2019. Disponível em <https://www.statisticssolutions.com/what-is-logistic-regression/>. Acesso em 02 set. 2019.

(50) SWISS TIMING. **Figure skating**. 2018. Disponível em <https://www.swisstiming.com/sports/figure-skating/>. Acesso em 20 set. 2019.

(51) THALER, S.L. Artificial neural networks. **Servo Magazine**. USA: T&L Publications, Apr. 2005.

(52) US FIGURE SKATING. **Identifying Jumps**. 2019. Disponível em <https://www.usfigureskating.org/content/Identifying%20Jumps.pdf>. Acesso em 27 set. 2019.

(53) WHITLEY, D.; SUTTON, A. M. Genetic Algorithms – A survey of models and methods. In: ROZENBERG, G.; BÄCK, T.; KOK, J. N. **Handbook of Natural Computing**. Berlin, DEU: Springer, 2012, p. 637-671.

(54) XU, C. ET AL. Learning to score figure skating sport videos. **Journal IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)**. IEEE, Early access, 2019.

(55) XU, J. **An Intuitive Guide to Deep Network Architectures**. 2017. Disponível em < https://towardsdatascience.com/an-intuitive-guide-to-deep-network-architectures-65fdc477db41>. Acesso em 15 jun. 2019.

(56) YADAV, J. S.; YADAV, M.; JAIN, A. Artificial neural network. **International Journal of Scientific Research and Education**. IND: Ijsae, v. 1, n. 6, 108-118, 2013.

(57) YAMATO, J.; OHYA, J.; ISHII, K. Recognizing human action in time-sequential images using hidden Markov model. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 1992, Champaign. **Proceedings**… Champaign, USA, 1992. p. 379-385.

(58) YI, F.; MOON, I. Image segmentation: a survey of graph-cut methods. In: INTERNATIONAL CONFERENCE ON SYSTEMS AND INFORMATICS (ICSAI), 2012, Yantai. **Proceedings**… Yantai, CHN, 2012. p. 1936-1941.

(59) YU, L.; LIU, H. Efficient Feature Selection via Analysis of Relevance and Redundancy. **The Journal of Machine Learning Research**. JMRL, v. 5, 1205-1224, 2004.

(60) ZHANG, S.; ZHANG, C.; YANG, Q. Data Preparation for Data Mining. **Journal Applied Artificial Intelligence**. Taylor & Francis, v. 17, 375-381, 2003.