UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

Victor Hugo Braguim Canto

Identificação Biométrica de Animais Baseada em Aprendizado de Máquina

Bauru
2023

Victor Hugo Braguim Canto

Identificação Biométrica de Animais Baseada em Aprendizado de Máquina

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Estadual Paulista "Júlio de Mesquita Filho".

Orientador: Prof. Dr. Aparecido Nilceu Marana

Coorientador: Dr. Gustavo Botelho de Souza

Bauru
2023

Canto, Victor Hugo Braguim.
Identificação Biométrica de Animais Baseada em Aprendizado de Máquina / Victor Hugo Braguim Canto. – Bauru, 2023
65 f. : il., tabs.

Orientador: Prof. Dr. Aparecido Nilceu Marana
Dissertação (mestrado) - Universidade Estadual Paulista “Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Reconhecimento Facial de Cães. 2. Aprendizado de Máquina. 3. Aprendizado em Profundidade. I. Marana, Aparecido Nilceu II. Universidade Estadual Paulista “Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas. III. Identificação Biométrica de Animais Baseada em Aprendizado de Máquina.

CDU – 518.72:76

Ficha catalográfica elaborada pela biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

UNIVERSIDADE ESTADUAL PAULISTA

Campus de Bauru

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE VICTOR HUGO BRAGUIM CANTO, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO, DA FACULDADE DE CIÊNCIAS - CÂMPUS DE BAURU.

Aos 03 dias do mês de agosto do ano de 2023, às 09:00 horas, por meio de Videoconferência, realizou-se a defesa de DISSERTAÇÃO DE MESTRADO de VICTOR HUGO BRAGUIM CANTO, intitulada **Identificação Biométrica de Animais Baseada em Aprendizado de Máquina.** A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. APARECIDO NILCEU MARANA (Orientador - Participação Virtual), do Departamento de Computação, da Faculdade de Ciências, UNESP, campus de Bauru, SP, Profa. Dra. SIMONE DAS GRAÇAS DOMINGUES PRADO (Participação Virtual) do Departamento de Computação, da Faculdade de Ciências, UNESP, campus de Bauru, SP, e Dr. MARCUS DE ASSIS ANGELONI (Participação Virtual) da Samsung Eletrônica da Amazônia, Campinas, SP, Brasil. Após a exposição pelo mestrando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final **APROVADO**. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo Presidente da Comissão Examinadora.

gov.br
Documento assinado digitalmente
APARECIDO NILCEU MARANA
Data: 03/08/2023 14:05:52-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. APARECIDO NILCEU MARANA

Victor Hugo Braguim Canto

Identificação Biométrica de Animais Baseada em Aprendizado de Máquina

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Estadual Paulista "Júlio de Mesquita Filho".

Comissão Examinadora

Prof. Dr. Aparecido Nilceu Marana
UNESP - Câmpus de Bauru
Orientador

Dr. Marcus de Assis Angeloni
Samsung Eletrônica da Amazônia

Profa. Dra. Simone das Graças Domingues Prado
UNESP - Câmpus de Bauru

Bauru
3 de Agosto de 2023

Dedico este trabalho ao meu avô Paulo Aparecido Braguim (in memoriam), meu segundo pai! Obrigado por sempre estar ao meu lado! Um dia nos reencontraremos!

Agradecimentos

Primeiramente, gostaria de agradecer a Deus pelo dom da vida, pela força em todos os desafios e pela oportunidade em chegar em lugares que eu nunca imaginei que fosse possível.

Agradeço também à minha família, meu pai Amarildo, minha mãe Cleia, minha irmã Bruna e a minha avó Dercy. Obrigado por todo amor, apoio e suporte! Sem vocês nada disso seria possível!

À minha namorada Jaqueline, por todo amor, companheirismo, apoio, paciência e suporte. Te amo!

Ao meu orientador, Prof. Nilceu, por todo o aprendizado, incentivo e, principalmente por ter sido um excelente amigo nessa jornada. Muito obrigado, professor! Serei eternamente grato!

Ao meu coorientador, Gustavo, muito obrigado por toda ajuda e incentivo! Obrigado por aceitar coorientar este trabalho!

Ao João Renato, pelo auxílio na execução de experimentos e, também, na parceria no desenvolvimento do artigo publicado no Bracis.

Aos professores por todas as lições e ensinamentos que servirão pela vida toda.

À UNESP e todos os seus servidores pela oportunidade da oferta de um ensino de qualidade e por toda a estrutura que passei desde a graduação e, agora, no mestrado. Fez toda diferença na minha vida! Obrigado por tudo!

Também, a todos meus amigos que estavam comigo nessa jornada!

"Que os vossos esforços desafiem as impossibilidades, lembrai-vos de que as grandes coisas do homem foram conquistadas do que parecia impossível."

Charles Chaplin

Resumo

Nas últimas décadas, ocorreu uma popularização do uso de características biométricas para a identificação humana. Neste período, foi demonstrado por inúmeras pesquisas que a identificação biométrica é eficaz, eficiente e, apesar de também poder sofrer ataques, é mais segura do que as formas tradicionais de identificação baseada em posses e conhecimento. Recentemente, ocorreu também um aumento expressivo na demanda por métodos eficazes, eficientes e seguros para a identificação animal, devido à necessidade de rastreabilidade, gerenciamento e controle desta população, que cresce em taxas maiores do que a população humana, particularmente os animais domésticos de estimação. A identificação animal tem sido baseada na posse de dispositivos como placas em coleiras, *tags* e *chips* implantados, o que possibilita a ocorrência de muitas fraudes nos sistemas de identificação. Além disso, essa forma de identificação tem uma baixa cobertura populacional pois nem todos os indivíduos portam tais dispositivos. Diante desses fatos, a Biometria tem sido proposta também como uma forma mais adequada para a identificação animal, entretanto, ainda são escassos, na literatura, os trabalhos com foco na identificação biométrica animal. Esta dissertação de mestrado investiga e propõe métodos para a identificação biométrica de animais, com foco na identificação de cães, por meio de características biométricas faciais. De forma análoga ao que tem sido feito em trabalhos atuais para a identificação facial humana, neste trabalho, explorou-se abordagens baseadas no estado da arte em aprendizado de máquina em profundidade, como as Redes Neurais Convolucionais, mais especificamente a arquitetura *ResNet-50*, e os *Vision Transformers*, mais especificamente a arquitetura *EfficientFormer-L1*. Os métodos propostos foram avaliados sobre duas bases de dados, *DogFaceNet* e *DogID Dataset*, sendo a primeira uma base de dados pública contendo 8.363 imagens de 1.393 animais, que tem sido utilizada em trabalhos correlatos, e a segunda, uma base de dados bem maior, contendo 125.873 imagens de 39.148 animais, que foi desenvolvida neste trabalho. Os resultados obtidos sobre ambas as bases de dados mostraram que os métodos propostos foram exitosos para a identificação de cães, sendo que a arquitetura baseada em *Vision Transformers* superou consideravelmente a arquitetura baseada em Redes Neurais Convolucionais.

Palavras-chave: Biometria em Animais. Identificação de Cães. Reconhecimento Facial. Vision Transformers. Redes Neurais Convolucionais. ArcFace.

Abstract

In recent decades, there has been a popularization of the use of biometric characteristics for human identification. In this period, it has been demonstrated by numerous studies that biometric identification is effective, efficient and, although it can also be attacked, it is more secure than traditional forms of identification based on possessions and knowledge. Recently, there has also been a significant increase in the demand for effective, efficient and safe methods for animal identification, due to the need for traceability, management and control of this population, which grows at higher rates than the human population, particularly domestic pets. Animal identification has been based on the possession of devices such as plaques on collars, tags and implanted chips, which allows the occurrence of many frauds in identification systems. In addition, this form of identification has a low population coverage because not all individuals carry such devices. In view of these facts, Biometrics has also been proposed as a more adequate form for animal identification, however, studies focusing on animal biometric identification are still scarce in the literature. This master's thesis investigates and proposes methods for the biometric identification of animals, focusing on the identification of dogs through facial biometric characteristics. Analogously to what has been done in current works for human facial identification, in this work, approaches based on the state of the art in in-depth machine learning were explored, such as Convolutional Neural Networks, more specifically the ResNet-50 architecture, and the Vision Transformers, more specifically the EfficientFormer-L1 architecture. The proposed methods were evaluated on two databases, DogFaceNet and DogID Dataset, the first being a public database containing 8,363 images of 1,393 animals, which has been used in related works, and the second, a much larger database, containing 125,873 images of 39,148 animals, which was developed in this work. The results obtained on both databases showed that the proposed methods were successful for identifying dogs, and the architecture based on Vision Transformers considerably outperformed the architecture based on Convolutional Neural Networks.

Keywords: Biometrics in Animals. Identification of Dogs. Facial recognition. Visual Transformers. Convolutional Neural Networks. ArcFace.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

AI	*Artificial Intelligence*
ANN	*Artificial Neural Networks*
AP	*Average Precision*
AUC	*Area Under the Curve*
BARK	*Best Architecture to Retrieve K9*
BRISK	*Binary Robust Invariant Scalable Keypoints*
CMC	*Cumulative Match Characteristic*
CNN	*Convolutional Neural Network*
CPU	*Central Processing Unit*
EER	*Equal Error Rate*
GPU	*Graphics Process Unit*
HD	*Hard Disk*
IoT	Internet of Things
IoU	*Intersecction over Union*
IPB	Instituto Pet Brasil
mAP	*Mean Average Precision*
MB	*MetaTransformers Block*
MLP	*Multilayer Perceptron*
NLP	*Natural Language Processing*
NMS	*Non-Maximum Suppression*
OMS	*Organização Mundial de Saúde*
ORB	*Oriented FAST and Rotated BRIEF*
RAM	*Random-Access Memory*

ReLU	*Rectified Linear Unit*
ResNet	*Residual Network*
RFID	*Radio Frequency Identification*
RNN	*Recurrent Neural Network*
ROC	*Receiver Operating Characteristic*
SAM	*Segment Anything Model*
SIFT	*Scale-Invariant Feature Transform*
SURF	*Speeded Up Robust Features*
SVM	*Support Vector Machines*
USP	Universidade de São Paulo
VGG	*Visual Geometry Group*
ViT	*Vision Transformer*
YOLO	*You Only Look Once*

Sumário

1 Introdução

De acordo com o IPB (2021), em 2021 o Brasil possuía uma população de 149,6 milhões de animais de estimação (também chamados de *pets*), um aumento de 3,7% comparado ao ano de 2020. Dentre todos os animais de estimação, os cães estavam no topo da lista, com 58,1 milhões de indivíduos, seguidos pelas aves, com 41 milhões, e pelos gatos, com 27,1 milhões. Em percentual de crescimento comparado com o ano de 2020, os cães apresentaram um aumento de 4,5%, os gatos de 6% e as aves de 1,5%.

Outro ponto a se considerar é que os números não crescem apenas em relação à população de animais, mas também ao varejo *pet*. De acordo com o IPB (2022), em 2022, o mercado *pet* teve um faturamento de R$ 60,2 bilhões, um aumento de 16,4% comparado ao ano de 2021. Também, o segmento de serviços veterinários teve um aumento de 16,2%, clínicas e hospitais veterinários um aumento de 17,3% em relação ao ano de 2021.

Observa-se, portanto, com base nos dados apresentados anteriormente, que o segmento de serviços veterinários é bastante expressivo. Por isso, assim como ocorre na identificação de humanos, a identificação de animais também está sujeita a fraudes. Por exemplo, no Reino Unido, segundo Youtalk-Insurance (2018), os casos de fraude de seguros de animais estão aumentando 400% ao ano, o equivalente a 2 milhões de libras por ano em sinistros. Sendo assim, é cada vez mais necessário o uso de métodos robustos para a identificação animal de forma rápida, eficaz e segura, de tal modo a desestimular, dificultar e até mesmo evitar fraudes.

Outro problema importante nesta área, é a falta de controle populacional adequado de animais que impacta diretamente na saúde pública. Segundo o Jornal da USP (USP, 2021), dados divulgados pela Organização Mundial de Saúde (OMS) indicam que no Brasil há aproximadamente 30 milhões de animais perdidos ou abandonados, sendo 10 milhões de gatos e 20 milhões de cães. Com esta falta de controle, poderá ocorrer um aumento significativo da população desses animais, sendo um fator de risco, pois animais podem transmitir doenças e parasitas. Por isso, é essencial implementar métodos de identificação e monitoramento animal de forma eficiente. Um exemplo disso, ocorre na Coreia do Sul, onde, segundo Jang et al. (2020), todos os cães precisam ser submetidos, compulsoriamente, a métodos de identificação, sejam eles por dispositivos de RFID (*Radio Frequency Identification*), *tags pet* ou outros dispositivos. Segundo reportagem exibida recentemente por um programa de televisão[1], a Holanda conseguiu zerar o número de cachorros de rua. Atualmente, o país, que é conhecido por ser um dos mais *"pet-friendly"*, registra todos os cães por meio de microchips implantados nos animais. Com este dispositivo, sempre que um animal é encontrado nas ruas sem seu tutor, ele é rapidamente identificado e caso não possua o dispositivo implantado, é levado para um

[1] <https://g1.globo.com/fantastico/noticia/2023/07/09/holanda-consegue-zerar-o-numero-de-cachorros-de-rua-do-pais.ghtml>

abrigo onde permanecerá até ser adotado.

Segundo Kumar et al. (2018), os métodos para identificação animal podem ser classificados em abordagens invasivas e não invasivas. Exemplos de abordagens invasivas são: marcações na orelha, tatuagens na orelha e implante de *microchips* no corpo do animal. Exemplos de abordagens não invasivas são: dispositivo de RFID (*Radio Frequency Identification*), coleiras com GPS (*Global Positioning System*), dispositivos de Internet das Coisas (IoT), dispositivos *bluetooth*, além dos métodos baseados em catacterísticas biométricas dos animais, tais como face, íris, retina e focinhos.

É importante observar que os dispositivos que são integrados ao corpo para fins de identificação têm, em geral, vida útil limitada e podem ser perdidos com o tempo. Ademais, segundo Jang et al. (2020), o uso de *microchips* pode trazer efeitos indesejados e afetar o bem-estar do animal. Por outro lado, os dispositivos que são portados pelos animais, como as coleiras de identificação, podem ser perdidas e até mesmo usadas em outros animais, para o cometimento de fraudes, por exemplo. Portanto, os métodos de identificação baseados em biometria animal parecem ser uma opção mais vantajosa, uma vez que podem ser eficazes e eficientes, além de não requerem a aquisição de dispositivos, que introduzem mais custos ao processo de identificação, e podem ser importantes para desestimular ou até mesmo prevenir fraudes.

Com o avanço computacional e tecnológico, houve uma popularização do uso de sistemas biométricos para a identificação humana em várias aplicações do cotidiano, como por exemplo, para se efetuar transações bancárias via aplicativos dos celulares, para se acessar academias e estádios de futebol, para as eleições, dentre outras tantas. Além disso, por meio da literatura, verifica-se que, por mais que a identificação biométrica possa sofrer ataques, ela ainda é o meio mais seguro de identificação de pessoas, quando comparada aos métodos de identificação baseados em posses (cartões, chaves, documentos, etc) e conhecimento (senhas, dados pessoais, etc).

Esta dissertação de mestrado visa investigar o estado da arte relacionado à identificação biométrica animal e propor métodos robustos, eficazes e eficientes para tal, uma vez que com o uso de características biométricas é possível realizar uma identificação não invasiva, automatizada e duradoura dos animais (neste caso, levando-se em conta os indivíduos adultos). Ademais, com o uso da biometria, não há custos para aquisição ou desenvolvimento de dispositivos, uma vez que é possível realizar a coleta da maioria das características biométricas utilizando-se as câmeras fotográficas existentes nos *smartphones*, por exemplo, que estão bastante disseminados na sociedade.

1.1 Objetivos

Esta dissertação de mestrado visa investigar a identificação biométrica de animais, com ênfase na identificação de cães, por meio das suas características biométricas faciais, e propor métodos robustos e eficazes baseados em aprendizado de máquina, em particular nas Redes Neurais Convolucionais (CNN - Convolutional Neural Networks) e nos *Vision Transformers* (ViT), tendo em vista os bons resultados reportados na literatura por métodos baseados nestes modelos para a identificação biométrica facial humana.

1.2 Estrutura da Dissertação

Esta dissertação está organizada em seis capítulos:

Capítulo 1: Apresenta a introdução ao tema, a motivação da proposta e os objetivos da dissertação de mestrado;

Capítulo 2: Apresenta o conceito de Biometria, conceitos de Redes Neurais Convolucionais com o uso da *ResNet* e *YOLO*, conceito de *Vision Transformers* com o uso da arquitetura *EfficientFormer*, conceitos da ArcFace e conceitos do uso de *Autodistill* em Visão Computacional;

Capítulo 3: Apresenta os principais métodos correlatos encontradas na literatura para identificação biométrica em animais;

Capítulo 4: Apresenta o método proposto neste trabalho para a identificação de animais baseada em características faciais;

Capítulo 5: Apresenta os resultados obtidos nos experimentos realizados, bem como as bases de dados utilizadas. É importante ressaltar que uma das bases de dados utilizada nos experimentos, denominada *DogID Dataset*, foi desenvolvida neste trabalho, sendo esta uma das contribuições importantes desta dissertação de mestrado;

Capítulo 6: Apresenta as conclusões deste trabalho, direcionamentos para trabalhos futuros e a publicação realizada.

2 Fundamentação Teórica

Neste capítulo, são apresentadas as bases teóricas que contém conceitos de Biometria, Redes Neurais Convolucionais (CNN), com a abordagem das arquiteturas *ResNet* (*Residual Networks*) e YOLO (*You Only Look Once*) e, também, *Vision Transformers* (ViT), com o uso da arquitetura *EfficientFormer*.

2.1 Biometria e Reconhecimento Facial

Atualmente, há um crescimento significativo no uso de aplicações que realizam a identificação humana biométrica de maneira automatizada. Segundo Rathgeb, Pöppelmann e Gonzalez-Sosa (2020), o processo de reconhecimento é realizado com base nas características do indivíduo, sendo elas biológicas ou comportamentais. Segundo Zhang (2013), para servirem ao processo de identificação as características biométricas dos indivíduos devem atender às seguintes condições.

- **Universalidade:** Todo indivíduo deve possuir a característica;
- **Unicidade:** A característica de cada indivíduo deve ser distinta dos demais indivíduos;
- **Permanência:** A característica não pode mudar significativamente ao longo do tempo;
- **Coletabilidade:** A característica deve poder ser medida quantitativamente;
- **Aceitabilidade:** A população de indivíduos a serem identificados deve aceitar aquela forma de identificação;
- **Desempenho:** A característica deve possibilitar uma identificação eficaz e eficiente, e ser robusta a fatores tais como mudanças ambientais, sensores, ruídos, etc;
- **Fraude:** A característica deve ser difícil de ser fraudada e, assim, dificultar ataques ao sistema biométrico.

O reconhecimento de faces é a forma mais natural de identificação utilizada pelo homem, por isso possui uma alta aceitação (JAIN; LI, 2011). Aliado a isso, o reconhecimento de faces pode ser efetuado à distância sem a participação consciente e direta do indivíduo sendo identificado, diferentemente de outras características biométricas, como as impressões digitais e a íris, que exigem colaboração do indivíduo durante o processo de identificação. Esta particularidade desta característica biométrica a torna bastante propícia para a identificação biométrica animal.

O reconhecimento de faces requer duas etapas básicas iniciais fundamentais:

- **Detecção da Face:** Responsável por detectar a face na imagem de entrada do sistema;
- **Extração de Características:** Responsável por extrair os descritores da face, que serão utilizados na etapa posterior, de reconhecimento.

Apesar do reconhecimento facial 2D ter atingido um nível significativo de maturidade e uma elevada taxa de sucesso, ele continua sendo uma das áreas de investigação mais ativas em visão computacional (KORTLI et al., 2020). Segundo Kortli et al. (2020), destacam-se três técnicas particularmente promissoras para o aprofundamento dessa área: (i) o desenvolvimento de métodos de reconhecimento facial 3D; (ii) a utilização de métodos de fusão multimodal de tipos de dados complementares, em particular os baseados em imagens visíveis e infravermelhas; e (iii) a utilização de métodos de aprendizado de máquina profundo. Nesta dissertação, exploramos modelos estado da arte de aprendizado de máquina profundo para a identificação de animais por meio de suas características faciais.

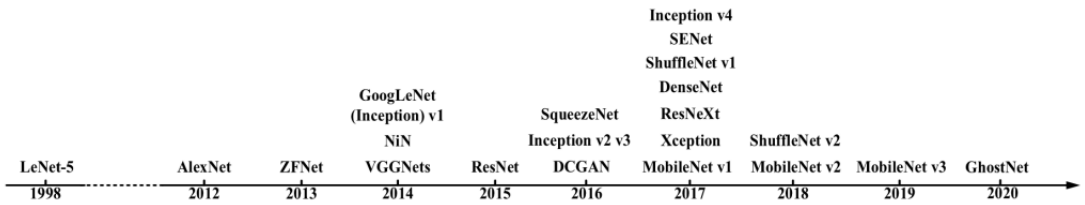
2.2 Redes Neurais Convolucionais

Nesta seção, são detalhados os conceitos e as arquiteturas de aprendizado de máquina em profundidade. Dessa forma, é apresentada uma introdução às Redes Neurais Convolucionais (CNN), à arquitetura *ResNet* e, também, à YOLO (*You Only Look Once*).

As Redes Neurais Convolucionais (CNN) são um tipo de Redes Neurais Artificiais (ANN) que, segundo Goodfellow, Bengio e Courville (2016), utilizam operações matemáticas denominadas convolução. Ainda, segundo Chollet (2016), as CNNs são amplamente utilizadas em problemas de imagens, como por exemplo, detecção de objetos, segmentação de imagens, reconhecimento facial, entre outros.

Com isso, segundo Chollet (2016), há algumas arquiteturas de CNNs com alta performance, como por exemplo, a *AlexNet*, VGG, *GoogleNet* e *ResNet*. Na Figura 1, há uma linha do tempo do surgimento de arquiteturas de CNNs.

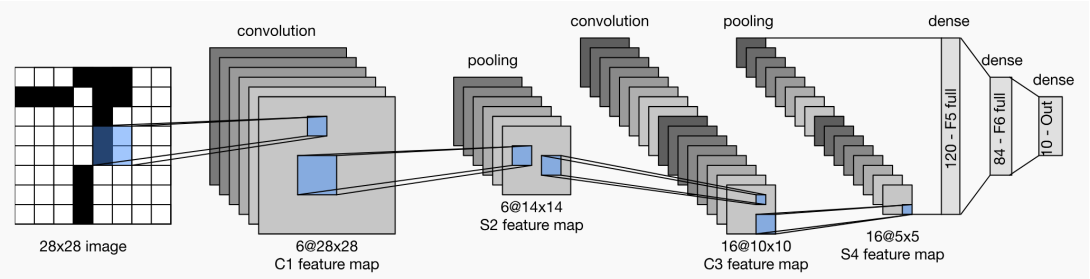
Figura 1 – Linha do tempo do surgimento das arquiteturas de CNNs.



Fonte: Adaptado de Li et al. (2021).

As Redes Neurais Convolucionais são organizadas em várias camadas, como por exemplo, camadas de entrada, as camadas convolucionais, camadas de *pooling* e camadas totalmente conectadas. A Figura 2 mostra a arquitetura da LeNet-5, uma CNN precursora, proposta por LeCun et al. (1998) para o reconhecimento de dígitos, que consiste em duas partes: um codificador convolucional contendo duas camadas convolucionais e um bloco denso contendo três camadas totalmente conectadas.

Figura 2 – Arquitetura da LeNet-5.



Fonte: <https://pt.d2l.ai/chapter_convolutional-neural-networks/lenet.html>

A camada convolucional é responsável por realizar operações de convolução nos dados de entrada de uma CNN. Com isso, segundo Goodfellow, Bengio e Courville (2016), a camada de convolução tem uma entrada, denominada pela função x e, o segundo argumento, a função ω sendo os filtros espaciais denominados de *kernels*. Assim, a operação de convolução, denotada pelo símbolo $*$, e a Equação 1 representam a operação, sendo sua saída s chamada mapa de características.

$$s(t) = (x * \omega)(t) \tag{1}$$

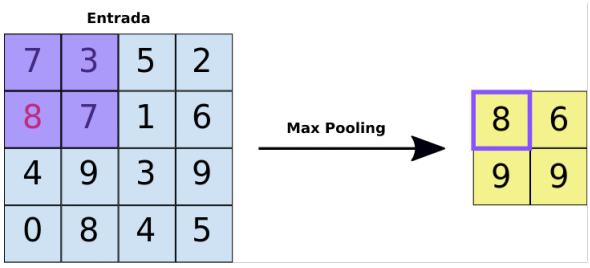
Por meio da camada de convolução, a arquitetura de uma CNN apresenta algumas vantagens, como por exemplo, melhora no custo computacional de execução, tratamento de variações de transformações geométricas como translações, redução de dimensionalidade e extração de características locais, como o caso de bordas e texturas. Segundo Kim (2017), além de realizar a extração de características com o uso da operação matemática de convolução, os ajustes dos pesos do algoritmo *backpropagation* são realizados nesta camada.

Após a aplicação de operações de convolução na camada convolucional, as funções de ativação, segundo Gu et al. (2018), são utilizadas para introduzir não-linearidades, limitar o intervalo do mapa de características e, consequentemente, com base na saída, haverá a definição da ativação de neurônios. Alguns exemplos mais comuns de funções de ativação são as funções ReLU, sigmóide e a tangente hiperbólica.

Na sequência, segundo Kim (2017), sobre os resultados gerados pelas camadas de convolução e após a aplicação da função de ativação, vem a camada de *pooling*. A camada de

pooling utiliza regiões de tamanho $n \times m$ para realizar uma combinação dos valores dos *pixels* vizinhos gerando um único valor. Com esse tipo de operação, é possível minimizar o tempo de execução no treinamento de arquiteturas de CNNs e, consequentemente, há a minimização do uso de recursos computacionais. Exemplos de funções de *pooling* são a *Max Pooling* e a *Average Pooling*. Na Figura 3, há o exemplo de uma operação de pooling, especificamente a operação de *Max Pooling* que, como saída, devolve o maior valor de cada vizinhança 2×2 da imagem de entrada.

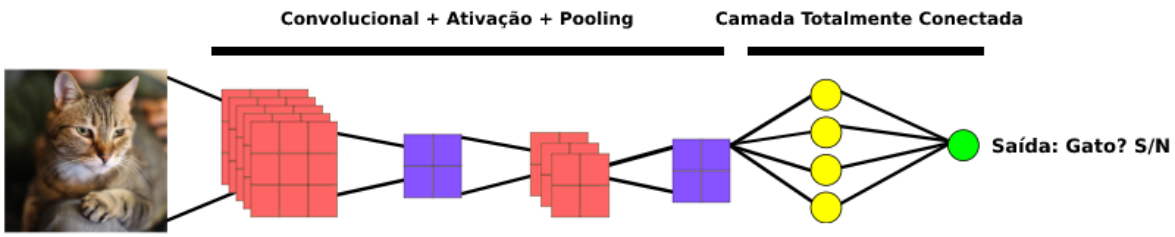
Figura 3 – Exemplo da operação de *Max Pooling* em uma vizinhança 2×2.



Fonte: Adaptado de GoogleDevelopers (2023).

Por fim, após a realização das etapas anteriores, há a camada totalmente conectada. A camada totalmente conectada é responsável por gerar o resultado final, seja ele, classificação, detecção, segmentação, entre outros. Em resumo, por exemplo, em uma rede com foco na verificação se há ou não um gato na imagem, a saída será uma sinalização binária da existência ou não existência de um gato na imagem. Na Figura 4, há um exemplo de uma CNN, com uma camada totalmente conectada ao final que efetua uma classificação binária para decidir se há ou não um gato na imagem de entrada.

Figura 4 – Exemplo de uma CNN contendo camadas de convolução, ativação, pooling e uma camada totalmente conectada ao final, para a classificação binária.



Fonte: Adaptado de GoogleDevelopers (2023).

2.2.1 *Residual Networks - ResNet*

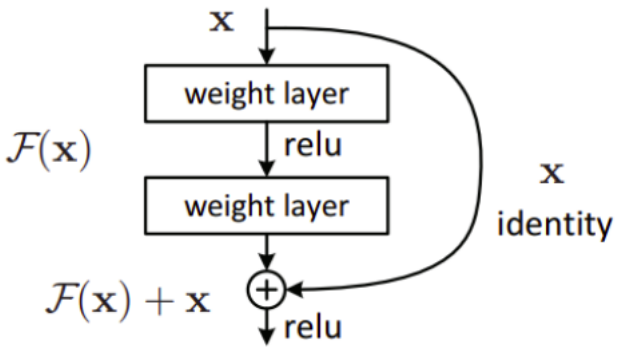
A *ResNet* foi proposta por He et al. (2016) e é um tipo de Rede Neural Convolucional. A arquitetura *ResNet* possui algumas variantes, como por exemplo *ResNet*-18, *ResNet*-34, *ResNet*-50 e *ResNet*-152.

A arquitetura de CNN *ResNet* propõe o uso de conexões de atalho durante o treinamento que auxilia no uso de informações para serem utilizadas em diversos pontos da rede e na resolução dos problemas de desempenho em CNNs tradicionais e, por isso, reduz o tempo de treinamento.

Nas arquiteturas *ResNet*, as conexões de atalho reduzem o problema da dissipação do gradiente, comum em CNN profundas. O problema da dissipação do gradiente é um fenômeno que ocorre durante o treinamento de redes neurais profundas, onde os gradientes usados para atualizar a rede tornam-se extremamente pequenos ou "desaparecem" à medida que são retropropagados das camadas de saída para as camadas anteriores.

Na Figura 5, há uma ilustração de uma conexão de atalho utilizada pela *ResNet* e na Figura 6, há o exemplo da arquitetura completa da *ResNet*-34.

Figura 5 – Conexão de atalho de uma *ResNet*.

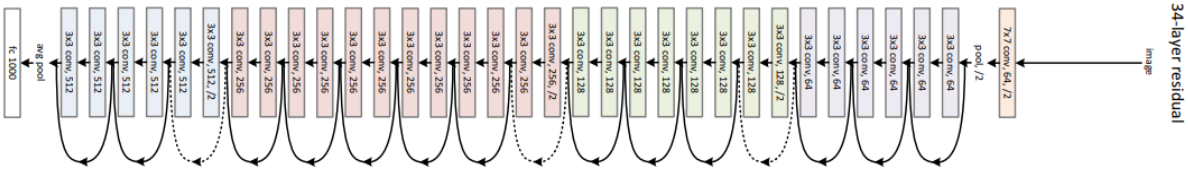


Fonte: GeeksforGeeks (2023).

Segundo Li et al. (2016), foram realizadas diversas melhorias na arquitetura da *ResNet* para aumentar o seu desempenho. Dentre elas, Zhang et al. (2017) propuseram conexões de atalho em vários níveis e Targ, Almeida e Lyman (2016) propuseram a utilização de mais camadas de convolução e fluxo de dados entre as camadas, entre outras.

Ainda, com o objetivo de realizar melhorias na arquitetura *ResNet*, segundo Zhang et al. (2017), foi sugerido o aumento do número de conexões de atalho por toda a rede e, também, segundo Targ, Almeida e Lyman (2016), o aumento de mais camadas de convolução, aumentando assim o fluxo de informações.

Figura 6 – Arquitetura completa da *ResNet*-34, em que é possível visualizar as várias conexões de atalhos existentes.



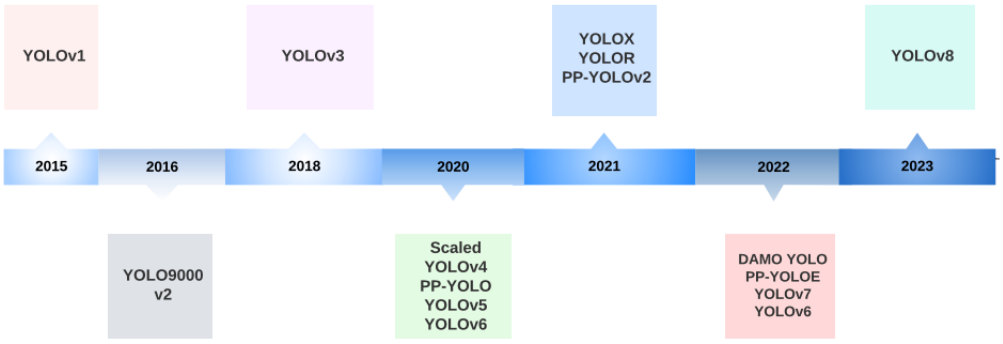
Fonte: Adaptado de GeeksforGeeks (2023).

2.2.2 YOLO

A arquitetura YOLO (*You Only Look Once*) proposta por Redmon et al. (2016) é amplamente utilizada na detecção de objetos. Segundo Terven e Cordova-Esparza (2023), esta arquitetura permite identificar e rastrear objetos com rapidez em diversas aplicações, como por exemplo, no uso em veículos autônomos, na agricultura, na medicina, na área de segurança, na área de robótica, na análise de imagens com drones, entre outras. Ainda, segundo Terven e Cordova-Esparza (2023), a arquitetura YOLO teve seu desenvolvimento iniciado em 2015 e, atualmente, em 2023, está na versão 8.

A Figura 7 mostra a linha do tempo da arquitetura YOLO.

Figura 7 – Linha do tempo da arquitetura YOLO.



Fonte: Adaptado de Terven e Cordova-Esparza (2023).

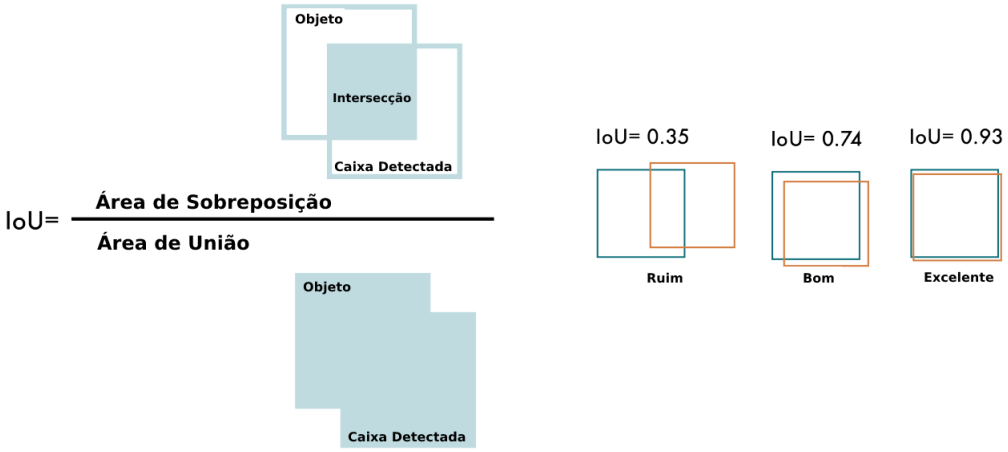
Em relação ao funcionamento da arquitetura YOLO, como na Figura 9, o item I é uma imagem de entrada que é dividida em uma grade de tamanho SxS. Logo após, no item II são realizadas as predições de cada elemento e, em seguida, no item III, é definida a probabilidade do elemento relativo à caixa delimitadora.

Depois, com a combinação dos itens II e III, com as caixas delimitadoras e a probabilidade relacionada a cada elemento, segundo Terven e Cordova-Esparza (2023), é aplicado o algoritmo

de Supressão Não-Máxima, do inglês *Non-Maximum Suppresion* (NMS), que possui como objetivo filtrar as caixas delimitadoras detectadas. Com isso, são geradas as saídas finais com o seu índice de confiança relacionado à cada classe.

Em relação à avaliação das predições, a arquitetura YOLO utiliza a média da precisão, do inglês *Average Precision* (AP) e também conhecida como *Mean Average Precision* (mAP) que é baseada em precisão e revocação. Segundo Terven e Cordova-Esparza (2023), baseado nessas métricas, o algoritmo define um valor para Intersecção sobre a União (IoU). A Figura 8 demonstra como é feita a classificação do valor da IoU.

Figura 8 – Definição da IoU - Ao lado esquerdo, a definição da área de intersecção e, ao lado direito, o valor e sua classificação.



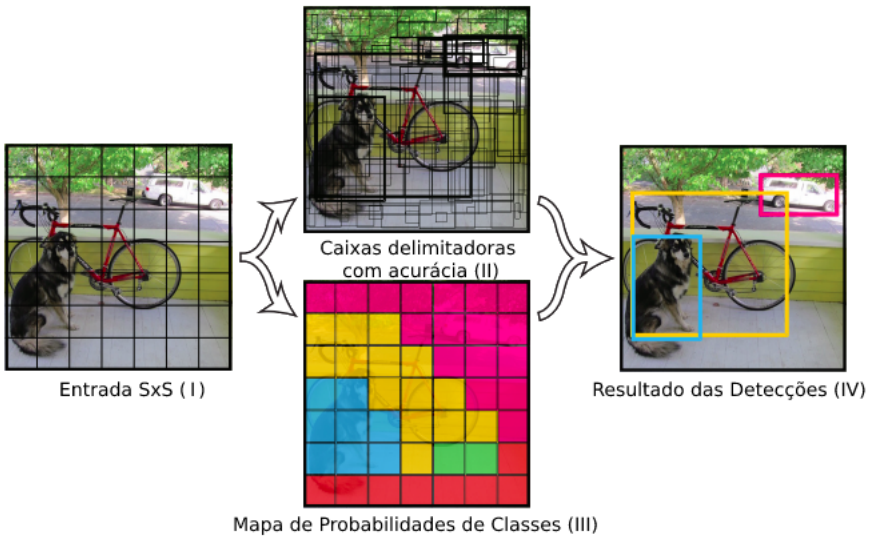
Fonte: Adaptado de Terven e Cordova-Esparza (2023).

2.3 *Vision Transformers* - ViT

Vision Transformer (ViT) é uma arquitetura de rede neural de aprendizado de máquina profundo desenvolvida para utilização em aplicações de Visão Computacional e inspirada nos excelentes resultados obtidos em arquiteturas de *Transformers* utilizadas em NLP (*Natural Language Processing*) (DOSOVITSKIY et al., 2020). O uso de *Transformers* foi proposto inicialmente por Vaswani et al. (2017) para tarefas de tradução de textos. Na Figura 10 há um exemplo da arquitetura de *Vision Transformer*.

Segundo Dosovitskiy et al. (2020), na primeira etapa, a imagem é dividida em *patches* de tamanho fixo, como ilustra a Figura 11. Após a divisão dos *patches*, ainda segundo Dosovitskiy et al. (2020), cada *patch* é transformado em um vetor unidimensional que é utilizado em uma camada de projeção linear que adiciona as posições e, na sequência, é enviado a um *Transformer Encoder*. Além disso, as posições são incorporadas para que durante o aprendizado tenha informações da localização espacial na imagem, pois conforme a Figura 12, caso a posição não

Figura 9 – Funcionamento da Arquitetura YOLO.



Fonte: Adaptado de Redmon et al. (2016).

seja informada, não seria possível uma interpretação correta dos elementos da imagem, pois há a ausência de recursos de memória, como aqueles utilizados em Redes Neurais Recorrentes (RNN).

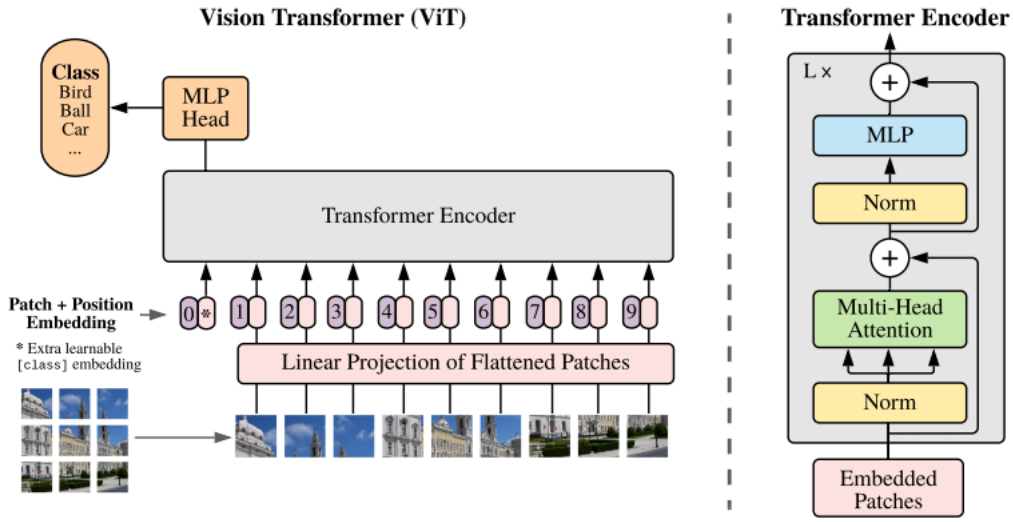
Ademais, o *Transformer Encoder* segue a mesma estrutura proposta por Vaswani et al. (2017) no uso em NLP. Com isso, na camada *Multi-Head Attention* há como entrada os *patches* transformados em vetores linearizados com suas posições. Assim, nesta fase, há o cálculo dos pesos de atenção de cada parte representada, indicando a sua importância. Por meio disso, segundo Vaswani et al. (2017), é gerado um conjunto de pares chave-valor na forma de vetores que são somados de forma ponderada permitindo a atenção nos *patches* mais importantes do processo.

Por fim, ainda segundo Vaswani et al. (2017), a saída da camada *Multi-Head Attention* é utilizada em uma camada de rede *Feed-Forward* totalmente conectada que aplica separadamente a cada posição. Além disso, há transformações lineares com a função de ativação ReLU. Sendo assim, com as saídas do *Transformer Encoder* é utilizada uma arquitetura *Multilayer Perceptron* (MLP) com o objetivo de gerar a saída do modelo.

2.3.1 Arquitetura *EfficientFormer*

A arquitetura *EfficientFormer* proposta por Li et al. (2022) é baseada em *Vision Transformers* propostos por Dosovitskiy et al. (2020). No trabalho proposto por Li et al. (2022) foi realizada uma análise de latência em dispositivos com o uso de *Vision Transformers*. Com isso, segundo Li et al. (2022) foram obtidas quatro observações:

Figura 10 – Arquitetura de uma Rede Neural ViT - *Vision Transformer*.



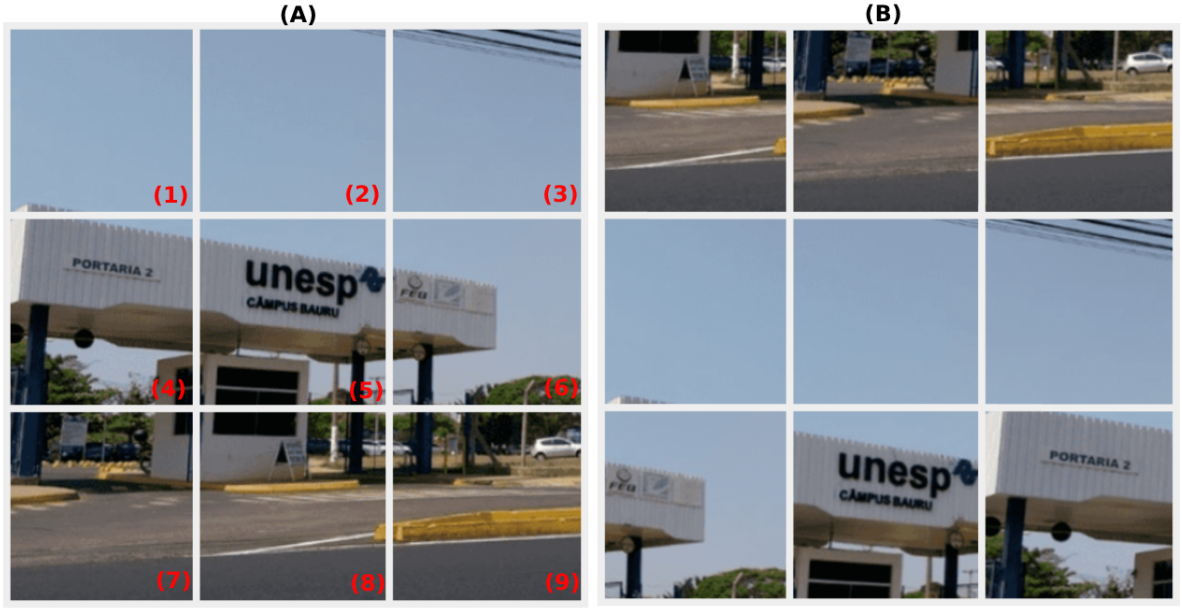
Fonte: Elaborado por Dosovitskiy et al. (2020).

Figura 11 – Divisão da imagem em *patches*.

Fonte: Elaborado pelo autor.

- A inclusão de *patches* grandes afeta o desempenho em dispositivos móveis;
- É importante a escolha do misturador de *tokens* (ou *patches* após a transformação em vetores com localização espacial na imagem) de arquiteturas de ViT, pois afetam diretamente no desempenho;
- O uso da implementação da arquitetura de *Multilayer Perceptron* (MLP) também é fator

Figura 12 – (A) Divisão da imagem em *patches* com posições incorporadas; (B) Divisão da imagem sem as posições conhecidas, impossibilitando a interpretação correta dos elementos.



Fonte: Elaborado pelo autor.

importante para o desempenho;

- A latência de execução depende da função de ativação, do *hardware* e do compilador do dispositivo.

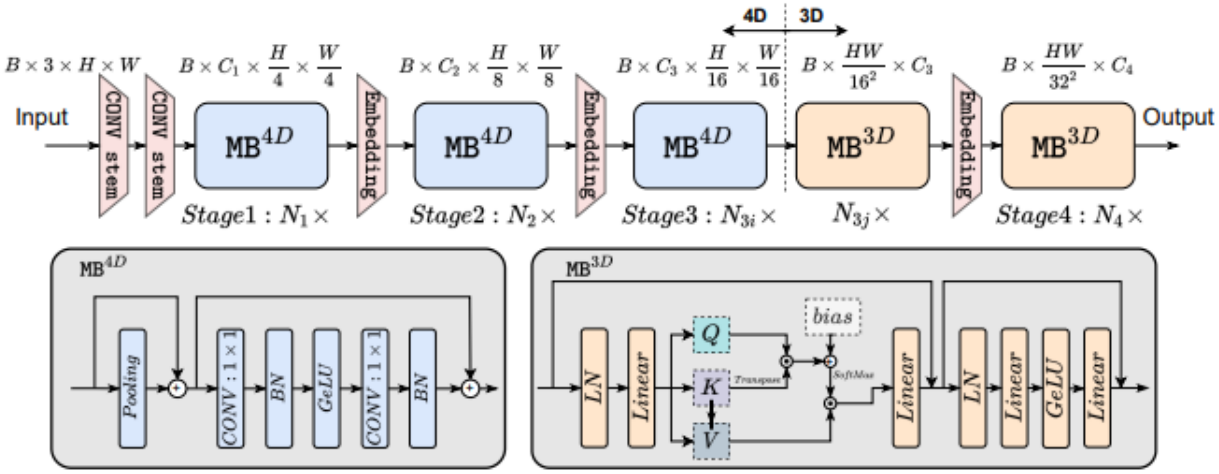
Dessa forma, após a análise de latência, foi proposta por Li et al. (2022) a arquitetura do *EfficientFormer* conforme a Figura 13.

Segundo Li et al. (2022), a arquitetura contém o uso de *patches* com uma pilha de blocos denominados *MetaTransformers Block* (MB), incluindo um misturador de *tokens* não especificado e seguido por bloco de *Multilayer Perceptron*. O objetivo é realizar a redução do *token* em cada etapa e, por consequência, realizar a redução da latência.

2.4 ArcFace

A ArcFace, segundo Deng et al. (2019), é uma função de erro que posiciona amostras de uma classe em uma hiperesfera de raio s de tal forma que as distâncias geodésicas entre os elementos sejam minimizadas e a distância ângular permaneça posicionada no intervalo m. A Equação 2 define a função de erro ArcFace.

Figura 13 – Arquitetura da *EfficientFormer*.



Fonte: Elaborado por Li et al. (2022).

$$L = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos(\theta_j)}} \tag{2}$$

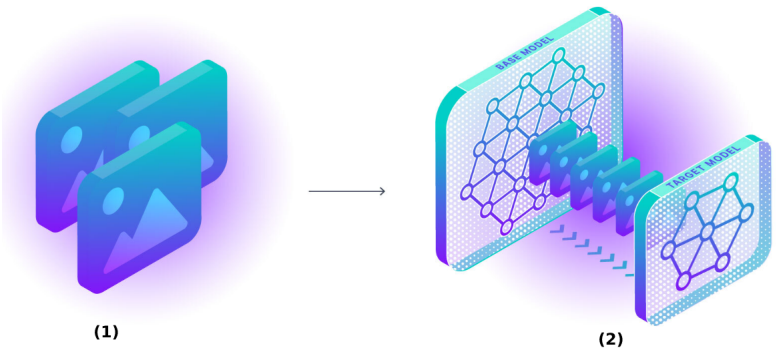
Ainda segundo Deng et al. (2020), é possível realizar uma análise em subespaços utilizando a função de erro ArcFace. Para isto, é necessária a adição de novos parâmetros, como por exemplo, o uso de subcentro para realizar a extração de recursos. O parâmetro adicional de subcentro é uma matriz de pesos com dimensões $C \times k$, onde C é o número de classes e k é o número de subespaços.

2.5 *Autodistill* em Visão Computacional

O processo de *Autodistill* em Visão Computacional, segundo Ultralytics (2023), é uma técnica que utiliza modelos maiores de base com o objetivo de construir modelos menores pré-treinados para realizar a detecção e segmentação de objetos ainda não rotulados. Dessa forma, por transferência de aprendizado, é possível a criação desse modelo menor e a realização da rotularização de bases. Na Figura 14, há um exemplo da arquitetura de funcionamento do processo de *Autodistill*.

Além disso, ainda segundo Ultralytics (2023), modelos muito grandes, como por exemplo o modelo lançado pela Meta AI, o SAM (*Segment Anything Model*), funcionando em tempo real, não trariam resultados significativos devido a sua alta dependência de hardwares, como o caso das Unidades de Processamento Gráfico (GPU). Assim, com o uso do processo de *Autodistill* é possível alocar esses modelos menores, mais especializados, em dispositivos com hardwares menos potentes, como por exemplo, em *smartphones*.

Figura 14 – Funcionamento da Arquitetura de Autodistill - Ao lado esquerdo, há a entrada de imagens não rotuladas e, ao lado direito, há o uso de modelos de base maiores treinados com milhões de imagens para a criação de modelos menores.

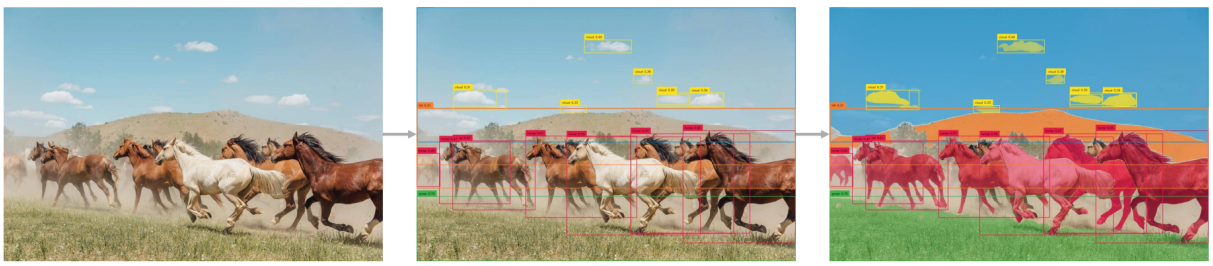


Fonte: Adaptado de Ultralytics (2023).

Ainda, segundo Ultralytics (2023), após definir a tarefa que será realizada pelo processo de *Autodistill*, a aplicação tem como entrada uma ontologia definindo qual tipo de objeto deverá ser detectado pelo novo modelo de aprendizado de máquina profundo.

A proposta do processo de *Autodistill* é baseada na combinação da utilização de detectores baseados em *Transformers* chamados de *GroundingDINO* e *GroundingSAM*. Segundo Liu et al. (2023), o objetivo é realizar a fusão entre modelos de NLP (*Natural Language Processing*) com Visão Computacional. Dessa forma, com uma *query* de textos, definida por uma ontologia, é possível realizar a detecção de objetos em imagens. Na Figura 15, há um exemplo dos textos de entrada e as detecções realizadas utilizando *Grounding DINO* e *Grounded-SAM*.

Figura 15 – Na entrada, há textos e uma imagem e na sequência há os dois processos de detecção: *Grounding DINO* e *Grounded-SAM*.



Fonte: Adaptado de Grounded-SAM (2023).

Assim, após a aplicação do processo de *Autodistill*, há como saída as posições dos

objetos detectados e a padronização para serem utilizados como entrada em modelos baseados na YOLO, ficando da seguinte forma: número da classe, posição x, y, w e h, sendo normalizados pela largura e altura da imagem, conforme as Equações 3:

$$
\begin{aligned}
x\_normalizado &= \frac{x}{largura\_imagem} \\
y\_normalizado &= \frac{y}{altura\_imagem} \\
w\_normalizado &= \frac{w}{largura\_imagem} \\
h\_normalizado &= \frac{h}{altura\_imagem}
\end{aligned}
\tag{3}
$$

3 Reconhecimento Biométrico em Animais

Neste capítulo, são apresentados trabalhos correlatos a este trabalho, que objetivam explorar técnicas de reconhecimento biométrico em animais. Algumas técnicas apresentadas são aplicadas em outros animais além de cães, o foco deste trabalho, o que permite em um futuro trabalho explorar também o seu uso.

3.1 Métodos de Reconhecimento de Focinhos

As informações do focinho podem ser utilizadas como característica biométrica de animais, sendo esta uma técnica de reconhecimento não invasiva.

Para coleta das informações do focinho podem ser utilizadas técnicas análogas à coleta da impressão digital humana, como, por exemplo, a impressão do focinho utilizando-se papel e tinta de carimbo ou utilizando-se sensores, como câmeras fotográficas digitais.

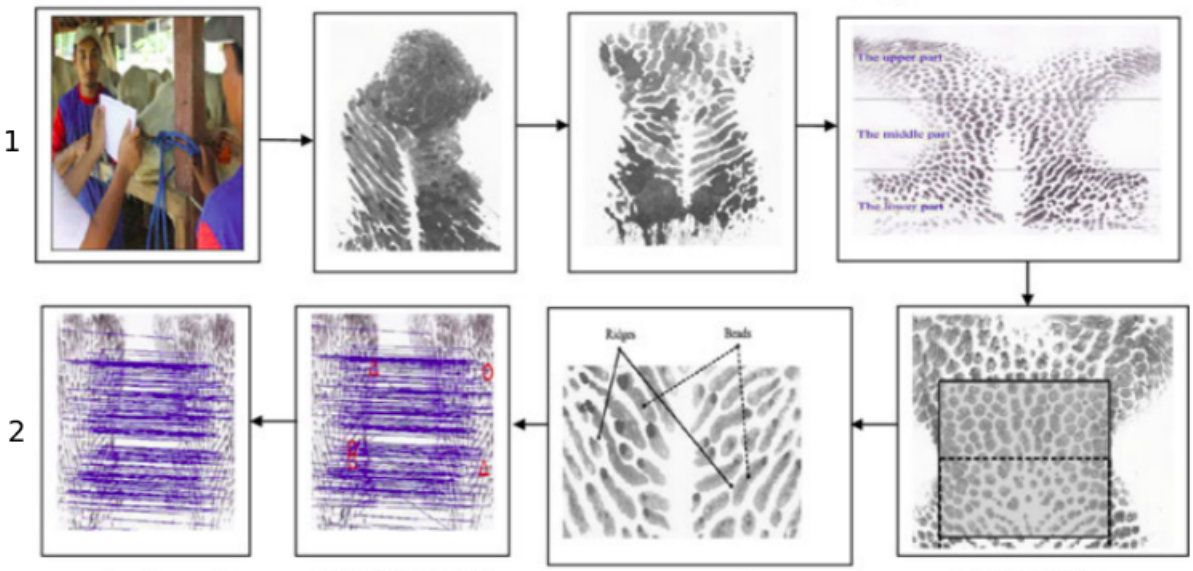
Kumar et al. (2018), por exemplo, propuseram um método com as seguintes etapas:

1. Imobilização do animal para coleta;
2. Limpeza do focinho;
3. Remoção de ruído do focinho;
4. Coleta da impressão do focinho;
5. Transformação da impressão coletada para imagem digital;
6. Pré-processamento e segmentação da imagem do focinho.

A Figura 16 ilustra o método proposto por Kumar et al. (2018). Ela mostra, na parte superior, as etapas do processo de coleta das impressões de focinhos de bovinos utilizando-se papel e tinta, seguida da digitalização da imagem coletada em uma imagem digital em tons de cinza e da segmentação da região de interesse e, na parte inferior, a etapa de extração de características e comparação das imagens.

Os estudos nesta área objetivam, em sua maioria, a identificação de bovinos. Entretanto, os padrões de focinhos de cães são similares aos de bovinos, com isso é possível realizar também o reconhecimento destes animais utilizando os traços biométricos dos focinhos. Jang et al. (2020) propõem o reconhecimento de cães utilizando câmeras como dispositivos para coleta das imagens dos focinhos. A Figura 17 mostra à esquerda um dispositivo utilizado para coleta

Figura 16 – Ilustração das etapas de coleta da imagem, extração de características e reconhecimento das impressões de focinhos de bovinos do método proposto por Kumar et al. (2018).



Fonte: Adaptado de Kumar et al. (2018)

de imagens de focinhos de cães e à direita uma imagem coletada, com a região de interesse para a identificação biométrica em destaque.

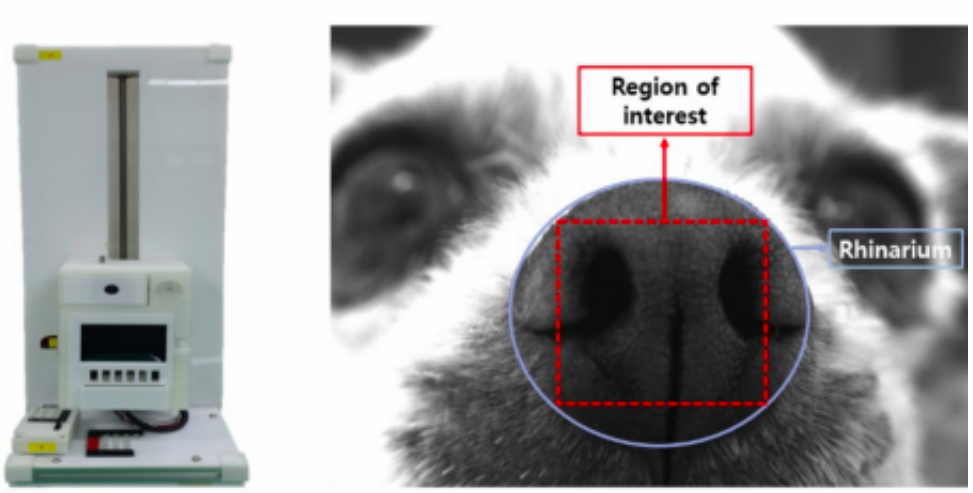
No método proposto por Jang et al. (2020), foram utilizados alguns algoritmos de extração de características para fins de comparação, como SIFT (*Scale-Invariant Feature Transform*), SURF (*Speeded Up Robust Features*), BRISK (*Binary Robust Invariant Scalable Keypoints*) e ORB (*Oriented FAST and Rotated BRIEF*). Para o SIFT e SURF, foram utilizados algoritmos não supervisionados para a busca de vizinhos mais próximos. Já para o BRISK e ORB, foi utilizada a distância de Hamming. Em relação aos resultados, o EER (*Equal Error Rate*) para o melhor algoritmo, ORB, foi de 0,35% para o processo de reconhecimento de focinhos.

3.2 Métodos de Reconhecimento de Retina

Segundo Kumar et al. (2018), o processo de reconhecimento de um animal utilizando os padrões de retina é semelhante ao realizado em humanos.

A Figura 18 mostra o processo de coleta do padrão de retina de um bovino e a Figura 19, nas imagens 1 e 2, mostra o padrão de retina coletado e, na imagem 3, mostra a extração de características do padrão de retina.

Figura 17 – Dispositivo utilizado para coleta de imagens de focinhos de cães (à esquerda) e a região de interesse de uma imagem coletada, utilizada para a extração de características (à direita).



Fonte: Jang et al. (2020)

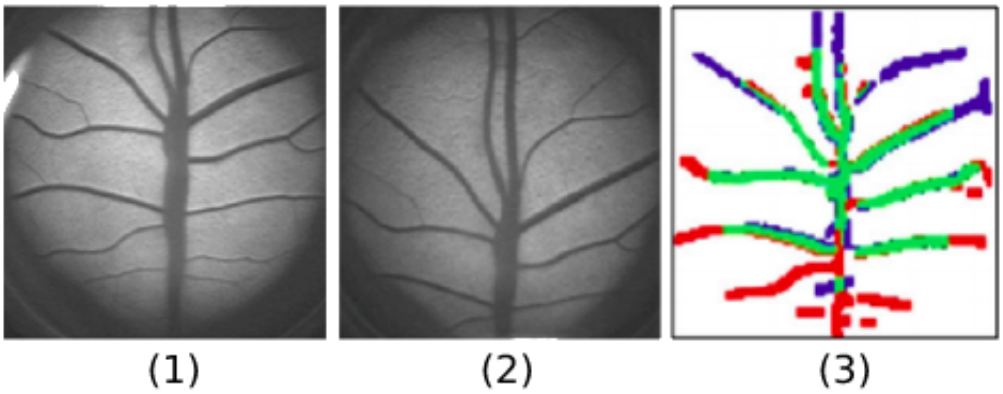
Figura 18 – Coleta de padrões de retina em um bovino.

Fonte: Kumar et al. (2018)

O estudo realizado por Kumar et al. (2018) coletou padrões de retina de 491 bois e 220 ovelhas e obteve uma acurácia de 96,2% para bovinos e 100% para ovinos.

Em um outro estudo, realizado por Allen et al. (2008), foram coletados padrões de retina em 1738 imagens de 869 bois e obteve-se uma acurácia de 98,30%.

Figura 19 – Exemplo de padrões de retina em bovinos.

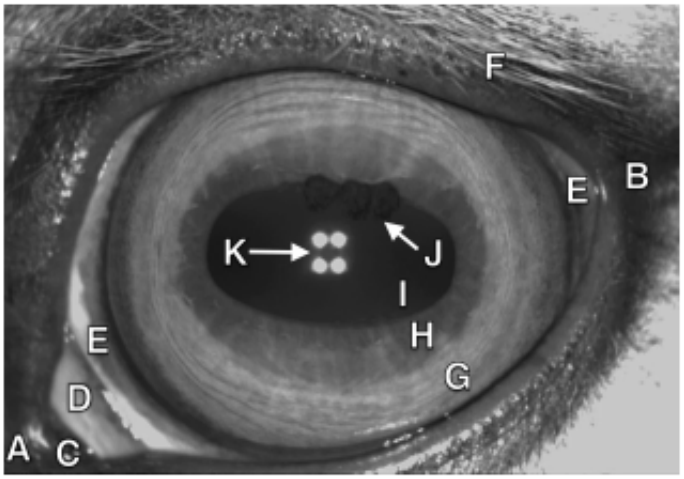


Fonte: Adaptado de Kumar et al. (2018)

3.3 Métodos de Reconhecimento da Íris e da Região Periocular

Segundo Trokielewicz e Szadkowski (2017), o método biométrico que utiliza a textura da íris, como proposto por Daugman (2009) para identificar pessoas, pode ser utilizado também para identificar animais, como por exemplo, cavalos. A Figura 20 mostra a imagem do olho de um cavalo capturada com uma câmera infravermelha, na qual é possível observar toda a estrutura ocular desses animais, incluindo as zonas ciliar e pupilar da íris.

Figura 20 – Estrutura Ocular de Equinos. A) Canto medial, B) Canto lateral, C) Carúncula lacrimal, D) Membrana livre nictitante, E) Esclera, F) Cílios, G) Zona ciliar da íris, H) Zona pupilar da íris, I) Pupila, J) Granula Iridica e K) Reflexão da fonte de luz infravermelha.



Fonte:Trokielewicz e Szadkowski (2017)

O método proposto por Trokielewicz e Szadkowski (2017) utilizou a fusão de caracte-

rísticas biométricas da íris e da região periocular e redes neurais profundas, obtendo uma taxa de EER (*Equal Error Rate*) de 9,5%.

3.4 Métodos Baseados em Reconhecimento Facial

Mougeot, Li e Jia (2019) propuseram o processo de verificação e identificação de cães para a base *DogFaceNet*. Foram utilizadas as arquiteturas profundas *VGG-like* e *ResNet-like*, sendo que a arquitetura que obteve o melhor desempenho foi a *ResNet-like*, com 92% em uma tarefa de verificação contendo 48 cães de um conjunto aberto de dados. Outrossim, no processo de identificação, também utilizando *ResNet-like* pelo melhor desempenho, os autores obtiveram 60,44% de acurácia para *rank-1* e 92% de acurácia para *rank-5*.

Um estudo realizado por Yoon, So e Rhee (2021), baseado nos estudos de Mougeot, Li e Jia (2019), propõe uma metodologia de utilização de espaços vetoriais para melhorar o desempenho no processo de identificação de cães. Dessa forma, no cenário de identificação facial de cães os autores propõem, usando *triplet loss*, a remoção da norma L2 para a utilização em um espaço vetorial além de uma esfera de raio 1. Com isso, os autores obtiveram uma acurácia de 97,33% em uma tarefa de verificação em um conjunto de dados aberto, apresentando um aumento de 5,79% na taxa de acurácia obtida por Mougeot, Li e Jia (2019).

O estudo conduzido por Moreira et al. (2017), propõe a utilização da base de dados *Flickr-dog* (e o seu complemento denominado *Snoopybook*). Foram utilizadas arquiteturas CNNs baseada em *OverFeat* e outra arquitetura baseada em redes otimizadas denominada BARK (*Best Architecture to Retrieve K9*). De maneira geral, a arquitetura baseada em *OverFeat* teve um resultado melhor (89,4%) comparado à arquitetura BARK (81,1%), sendo a diferença entre ambos de 8,3 pontos percentuais.

O estudo conduzido por Tu, Lai e Yanushkevich (2018) propõe o uso de duas abordagens para avaliar o processo de identificação na base de dados *Flickr Dog Dataset* com a arquitetura *DogNet*. Para a *DogNet* foi aplicada a transferência de aprendizado. Além disso, os experimentos foram divididos em 3 cenários: i) utilização de imagens somente de cães da raça *Pug*; ii) utilização de imagens somente de cães da raça *Husky* e; iii) utilização de imagens de cães de ambas as raças.

Dessa forma, comparado ao estudo de origem proposto por Moreira et al. (2017), a arquitetura *DogNet* teve melhor desempenho considerando a acurácia nos 3 experimentos, sendo que no cenário contendo apenas indivíduos da raça *Pug* a diferença foi de 21,8%, no cenário contendo apenas indivíduos da raça *Husky* a diferença foi de 14,91% e no caso de ambas as raças a diferença foi de 17,04%.

No estudo conduzido por Lai, Tu e Yanushkevich (2019), os autores propõem uma abordagem para fusão de características utilizando *soft biometrics*, no caso a raça, a altura

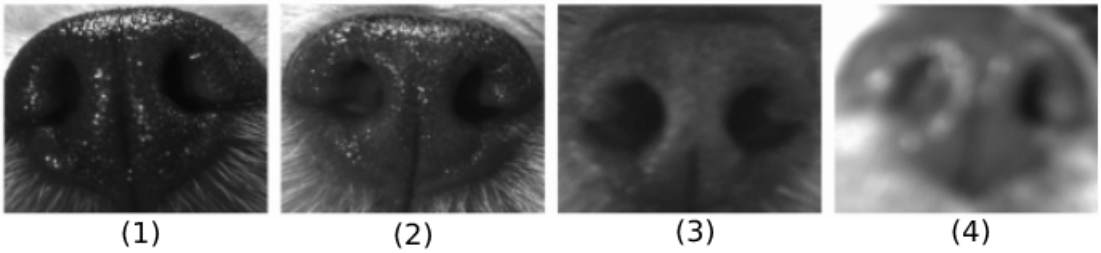
ou o gênero, além do uso de faces para identificação dos indivíduos. Para a classificação de raças, foi utilizada a base proposta por Moreira et al. (2017) e foram aplicadas as seguintes arquiteturas de CNNs: *Inception V3*, *MobileNet*, *VGG-16* e *Xception*. Neste estudo, de maneira geral, a arquitetura *MobileNet* teve melhor desempenho e, com isso, utilizando *soft biometrics* obteve uma acurácia de 90,80% e 91,29% no cenário onde há diferenciação das duas raças da base, ao passo que no cenário sem a utilização de *soft biometrics*, a acurácia foi de 84,94%.

3.5 Desafios no Reconhecimento Biométrico em Animais

As abordagens de reconhecimento de animais mencionadas anteriormente possuem algumas desvantagens, sejam elas durante o processo de coleta de dados ou, também, durante o processo de identificação do animal.

Em relação à abordagem de utilização de informações do focinho, na questão de coleta por meio de tinta e papel, exige-se uma precisão para manter o animal sem movimento e, também, necessita-se de uma boa higienização do focinho. Já na abordagem baseada na coleta dos focinhos por câmeras, é necessária uma câmera de alta resolução, higienização, minimização de inclinações e ter ausência de obstruções. A Figura 21 mostra experimentos realizados por Jang et al. (2020) de imagens de focinhos que foram descartadas, as imagens 1 e 2 por reflexão de luz e as imagens 3 e 4 por desfoque.

Figura 21 – Imagens descartadas de focinhos.



Fonte: Adaptado de Jang et al. (2020)

Na abordagem de utilização de faces, também é necessário ter ausência de obstruções, minimizar a inclinação e pose do animal, manter uma boa iluminação, evitar alterações de expressões do animal e evitar acessórios no rosto do animal (por exemplo, proteção do focinho). Além disso, uma grande dificuldade encontrada é em relação ao tempo transcorrido desde a captura da imagem facial do animal, pois podem haver mudanças significativas do animal. Outro desafio é a distinção entre indivíduos da mesma raça, como ilustra a Figura 22.

Figura 22 – Exemplos de indivíduos da mesma raça canina.

Fonte: Terra (2014).

Nas abordagens que utilizam informações dos olhos, é necessário evitar a movimentação do animal durante a coleta para que tenha precisão. Outro problema que pode ocorrer são doenças oculares que irão dificultar o processo de coleta e, consequentemente, de identificação.

4 Método Proposto

Este capítulo apresenta o método proposto neste trabalho para a identificação de animais utilizando aprendizado em profundidade.

4.1 Arquitetura do Sistema Biométrico

O sistema biométrico proposto para a identificação animal é composto por quatro módulos, a saber:

- **Módulo 1 - Detecção e Segmentação do Animal:** O módulo tem como entrada uma imagem e a saída será uma nova imagem com o animal segmentado, se houver animal;
- **Módulo 2 - Detecção e Segmentação da Face do Animal:** O módulo tem como entrada uma imagem e a saída será uma nova imagem com a face do animal segmentada, se houver face;
- **Módulo 3 - Extração de Características Biométricas Faciais do Animal:** O módulo tem como entrada a imagem resultante do processo de detecção e segmentação da face do animal (Módulo 2). Sua saída será um vetor de características da face do animal;
- **Módulo 4 - Identificação do Animal:** O módulo tem como entrada o vetor de características da face do cão e sua saída será a identidade do indivíduo.

4.1.1 Módulo 1 - Detecção e Segmentação do Animal

O Módulo 1 é responsável por realizar o processo de detecção e segmentação do animal. Assim, em outras palavras, verifica em uma imagem se há a presença ou não de cães e, caso haja, realiza a segmentação em uma nova imagem.

A Figura 23 mostra uma imagem de entrada contendo um cão, na sequência o módulo para detectar e segmentar o animal com a caixa delimitadora representada em vermelho e, por fim, a imagem recortada resultante, contendo apenas o animal.

Neste módulo, foi utilizada a arquitetura YOLO pré-treinada na sua versão 8 disponibilizada pela Ultralytics (2023).

Figura 23 – Ilustração do módulo de detecção e segmentação do animal.

Fonte: Elaborada pelo autor.

4.1.2 Módulo 2 - Detecção e Segmentação da Face do Animal

O Módulo 2 é responsável por realizar o processo de detecção e segmentação da face do animal. Este módulo verifica se há face de um cão na imagem. Caso haja, ele realiza a segmentação em uma nova imagem.

A Figura 24 mostra uma imagem de entrada contendo um animal, na sequência o módulo para detectar e segmentar faces com a caixa delimitadora da face representada em vermelho e, por fim, a imagem recortada resultante contendo apenas a face do animal.

Figura 24 – Ilustração do módulo de detecção e segmentação de face do animal.

Fonte: Elaborada pelo autor.

Neste módulo, foi utilizada a arquitetura YOLO na sua versão 8 treinada com dados

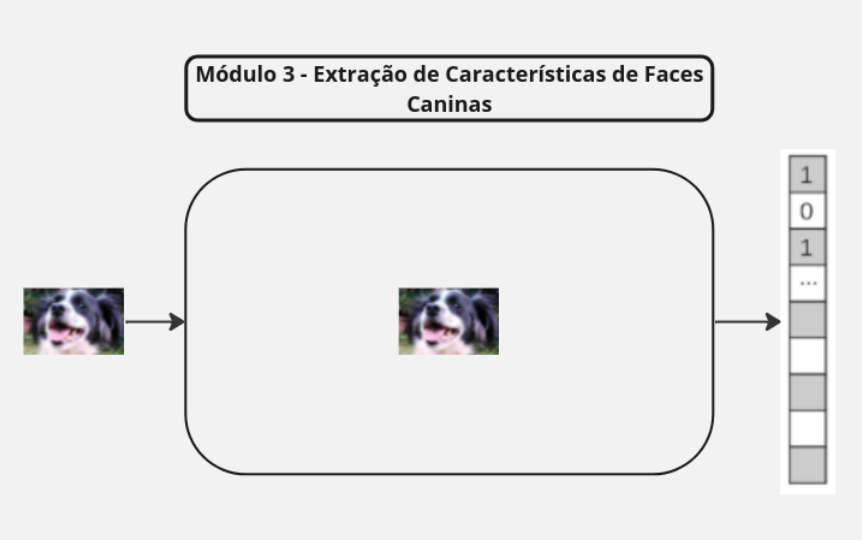
que foram rotulados por meio do processo de *Autodistill*.

4.1.3 Módulo 3 - Extração de Características Biométricas Faciais do Animal

O Módulo 3 é responsável por realizar o processo de extração de características biométricas faciais do animal. Neste caso, a entrada é uma imagem resultante do Módulo 2 e a saída é um vetor de características faciais do animal.

A Figura 25 ilustra a entrada e a saída do Módulo 3.

Figura 25 – Ilustração do módulo de extração de características biométricas faciais do animal.



Fonte: Elaborada pelo autor.

Neste módulo, foram utilizadas duas arquiteturas para extração de características faciais em cães, a *EfficientFormer-L1* e a *ResNet-50*.

4.1.4 Módulo 4 - Identificação do Animal

O Módulo 4 é responsável por realizar o processo de identificação do animal. A Figura 26 ilustra a entrada e a saída do módulo de identificação do animal.

Figura 26 – Ilustração do módulo de identificação biométrica do animal.

Fonte: Elaborada pelo autor.

5 Experimentos e Resultados

Neste capítulo são detalhadas as configurações de ambiente utilizadas, o processo da criação da base de dados, os experimentos realizados e são apresentados os resultados obtidos nos experimentos.

5.1 Configuração do Ambiente de Experimentos

Os experimentos foram realizados utilizando a linguagem de programação *Python* 3.9 e suas bibliotecas, como, Pytorch e Scikit-Learn. Além disso, em relação aos servidores, foram utilizadas as configurações apresentadas na Tabela 1:

Tabela 1 – Tabela de Configurações de Servidores

#	Versão Linux	Processador (vCPU)	Mem. RAM	Placa de Vídeo	HD
1	Ubuntu 20.04	Intel Xeon @ 8x 2.4 GHz	24 GB	2x NVIDIA GEFORCE RTX 3080 12 GB	90 GB (SSD)
2	Ubuntu 20.04	Intel Xeon @ 8x 2.4 GHz	24 GB	1x NVIDIA GEFORCE RTX 3090 24 GB	90 GB (SSD)

5.2 Bases de Dados

Nesta seção, serão detalhadas as bases de dados disponíveis na literatura com foco no reconhecimento biométrico de cães. Além disso, em relação à busca de bases de dados disponíveis publicamente, a Tabela 2 apresenta detalhes das palavras-chave (em Inglês e Português) e o número de bases de dados disponíveis publicamente encontradas a partir delas, considerando-se o *Google Scholar*, *Google* e *Kaggle* como fontes de busca.

5.2.1 *DogID Dataset*

A base de dados *DogID Dataset* foi proposta e desenvolvida neste trabalho, que será disponibilizada publicamente com fins exclusivamente não comerciais. A base contém 125.873 imagens de 39.148 cães, sendo pelo menos 2 imagens para cada cão. A Figura 27 mostra quatro exemplos de imagens de quatro cães, que compõem a base *DogID Dataset*.

Vale ressaltar também que, a base foi desenvolvida por meio do uso de *Webcrawlers* no site *Petfinder*[2], um site de adoção muito utilizado na América do Norte. Ainda, as imagens são registradas por usuários com interesses em divulgar animais disponíveis para adoção e, dessa forma, não há padronização em poses dos cães, iluminação das imagens, etc.

[2] <https://www.petfinder.com/>

Tabela 2 – Resultado das buscas de bases de dados para o reconhecimento facial canino, as fontes pesquisadas e as palavras-chave utilizadas.

Origem da Busca	**Palavras-chave utilizando operador *OR***	**Bases de Dados Disponíveis Publicamente**
Google Scholar (Em Inglês)	Dog Face Recognition Dog Face Identification Dog Facial Biometrics Canine Facial Biometrics Dog Face Recognition Dataset Dog Face Identification Dataset Dog Facial Biometrics Dataset Canine Facial Biometrics Dataset	*DogFaceNet* e *Flickr-dog Dataset*
Google Scholar (Em Português)	Reconhecimento Facial de Cães Identificação Facial de Cães Identificação Facial de Cães Biometria Facial de Cães Biometria Facial Canina Base de Dados de Reconhecimento Facial de Cães Base de Dados de Identificação Facial de Cães Base de Dados de Biometria Facial de Cães Base de Dados de Biometria Facial Canina	-
Kaggle	Dog Face Recognition Dog Face Identification Dog Facial Biometrics Canine Facial Biometrics Dog Face Recognition Dataset Dog Face Identification Dataset Dog Facial Biometrics Dataset Canine Facial Biometrics Dataset	-
Google (Em Inglês)	Dog Face Recognition Dog Face Identification Dog Facial Biometrics Canine Facial Biometrics Dog Face Recognition Dataset Dog Face Identification Dataset Dog Facial Biometrics Dataset Canine Facial Biometrics Dataset	*DogFaceNet* e *Flickr-dog Dataset*
Google (Em Português)	Reconhecimento Facial de Cães Identificação Facial de Cães Identificação Facial de Cães Biometria Facial de Cães Biometria Facial Canina Base de Dados de Reconhecimento Facial de Cães Base de Dados de Identificação Facial de Cães Base de Dados de Biometria Facial de Cães Base de Dados de Biometria Facial Canina	-

Vale ressaltar também que foi desenvolvida uma aplicação *client* em *Python* disponilizada em código aberto no GitHub[3] com instruções de uso para realizar a coleta da mesma base de dados diretamente no site da *Petfinder*. Caso não for realizado o uso diretamente pelo código disponibilizado, haverá um formulário com o preenchimento do solicitante e declaração de objetivos de uso para disponibilização de um link com a base de dados utilizada neste trabalho.

Além das imagens de cães, também há informações de raças, idade, porte, cores e gênero. Ainda, como resultado deste trabalho, também estão disponíveis as informações de *bounding boxes* do corpo inteiro dos cães e de suas respectivas faces.

[3] <https://github.com/vbraguimcanto/DogID_Dataset>

Figura 27 – Amostras de imagens da base de dados *DogID Dataset* desenvolvida neste trabalho.

Fonte: Elaborada pelo autor.

Outrossim, a base de dados possui cães de 224 raças e uma categoria de raças mistas. A Figura 28 mostra como a base está distribuída entre as raças, considerando-se as *top*-25 com mais indivíduos. Ainda, a Figura 29 apresenta a distribuição por sexo dos cães da base *DogID Dataset*. Em relação à idade dos cães, elas são divididas em adultos, jovens, bebês e sêniores. A Figura 30 mostra a distribuição pela idade dos cães da base *DogID Dataset*.

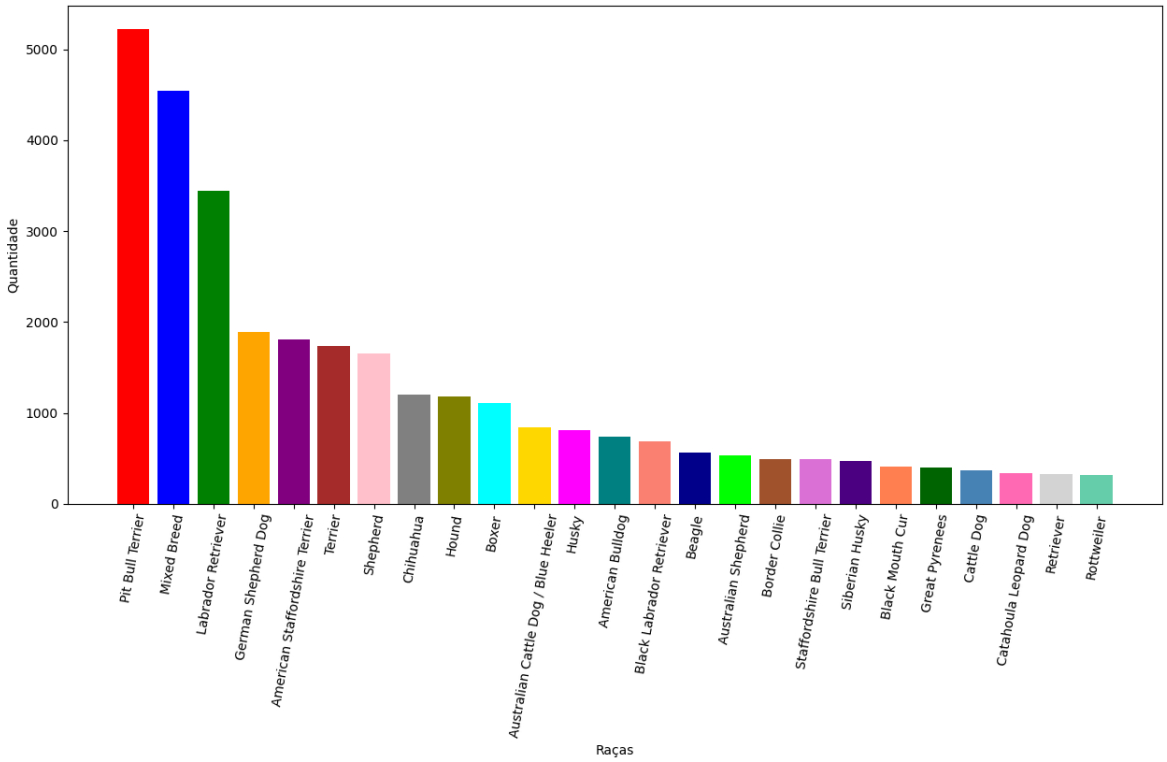
5.2.2 DogFaceNet

DogFaceNet é uma base de dados de imagens desenvolvida por Mougeot, Li e Jia (2019), contendo 8.363 imagens de 1.393 cachorros, sendo pelo menos 2 imagens por cachorro. A Figura 31 mostra imagens de quatro cães distintos da base de dados *DogFaceNet*.

5.2.3 Flickr-dog Dataset

Base de dados proposta por Moreira et al. (2017), contendo 374 imagens de 42 cães de apenas duas raças: *Pug* e *Husky*. A Figura 32 mostra imagens de quatro cães distintos da

Figura 28 – Histograma das *top*-25 raças de cães com maior número de indivíduos na base de dados *DogID Dataset*.



Fonte: Elaborada pelo autor.

base de dados *Flickr-dog Dataset*.

5.2.4 Snoopybook Dataset

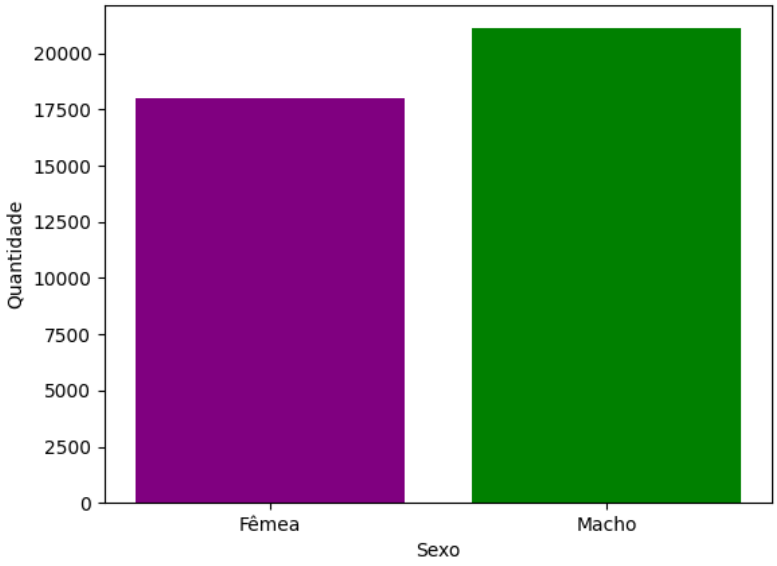
Esta base de dados é uma versão atualizada da base de dados *Flickr-dog Dataset*, na qual o autor adicionou 251 imagens de 18 cães mestiços. A Figura 33 mostra imagens de dois cães distintos da base de dados *Snoopybook Dataset*.

5.3 Detecção Facial de Cães

Para a realização do processo de detecção facial de cães, foi utilizado o YOLO na sua versão 8, lançada agora em 2023 pela Ultralytics (2023). YOLO é uma abreviatura para o termo "*You Only Look Once*". Trata-se de um algoritmo que detecta e reconhece vários objetos em uma imagem (em tempo real). A detecção de objetos no YOLO é feita como um problema de regressão e fornece as probabilidades de classe das imagens detectadas.

Nos experimentos para a detecção facial dos cães foi utilizada apenas a base de dados desenvolvida neste trabalho, a *DogID Dataset*, visto que a base de dados *DogFaceNet*

Figura 29 – Histograma dos sexos (macho e fêmea) dos cães da base *DogID Dataset*.



Fonte: Elaborada pelo autor.

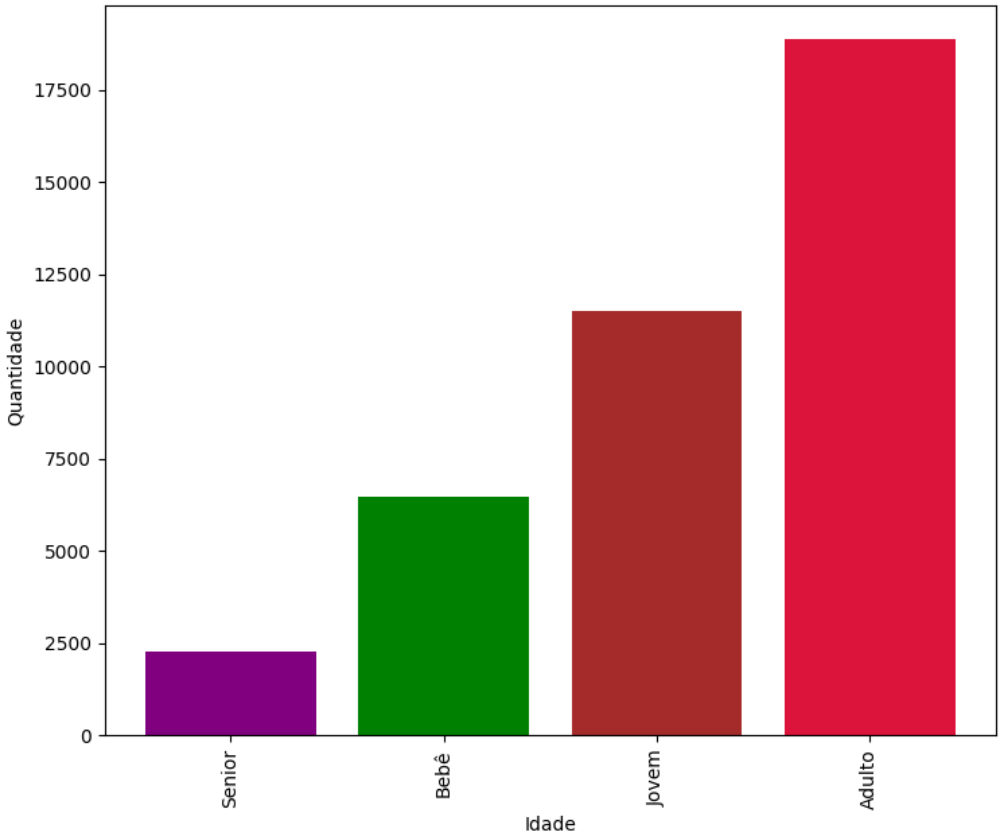
disponibiliza publicamente as imagens contendo as faces dos cães já detectadas e recortadas.

Com relação à base de dados *DogID Dataset*, nos dados coletados pelo *Webcrawler* não havia informações das posições dos cães e nem das suas faces nas imagens. Assim, por meio do *Autodistill*, foi realizada a rotularização da base de dados, seguindo as ontologias: *Dog Face*, *Canine Face*, *Dog Eyes* e *Dog Muzzle*. A Figura 34, mostra exemplos dos resultados da aplicação do *Autodistill* em algumas imagens da base de dados *DogID Dataset*, para a rotulação da base de dados e a detecção das faces dos cães nas imagens.

A arquitetura proposta para o sistema de detecção, contém o processo de detecção de cães em imagens e, logo após, detecção de faces em imagens, como ilustrado na Figura 35.

Após o processo de *Autodistill* foi utilizado o *threshold* de 0.8 em toda a base de dados, sendo a detecção em funil, iniciando na detecção de cães em imagens e, na sequência, a detecção de faces de cães em imagens. Além disso, imagens com detecção acima de 1 cão por imagem foram descartadas, sendo que a detecção de cães em imagens foram realizadas por meio de uma YOLO já pré-treinada. Assim, obteve-se uma acurácia de 87,2% nos dados em uma checagem realizada manualmente, para verificar a imagem original comparada com a predições geradas. Nos casos de erros de resultados na rotulação automática, os rótulos das imagens foram atribuídos manualmente.

Figura 30 – Histograma das idades dos cães da base de dados *DogID Dataset*, sendo bebê um cão com até 1 ano idade, jovem de 1 ano a 3 anos, adulto de 3 a 8 anos e sênior acima de 8 anos.



Fonte: Elaborada pelo autor.

5.4 Reconhecimento Facial de Cães

Para a realização do processo de reconhecimento facial de cães, foram utilizadas duas bases de dados: *DogFaceNet* e a *DogID dataset*, desenvolvida neste trabalho.

5.4.1 Experimentos na Base de Dados *DogFaceNet*

Conforme mencionado no Capítulo 3, Mougeot, Li e Jia (2019) propuseram uma base pública de imagens com foco no reconhecimento facial biométrico de cães, a *DogFaceNet*. A base original proposta no trabalho em questão possuía 3.148 imagens de 485 cães. Entretanto, novas imagens de cães foram adicionadas à base de dados e, com isso, ela passou a contar com 8.363 imagens de 1.393 cães, sendo pelo menos duas imagens por cão. Portanto, os experimentos realizados, e os respectivos resultados reportados nesta dissertação, não são totalmente aderentes e comparáveis diretamente aos realizados e reportados por Mougeot, Li e

Figura 31 – Imagens de quatro cães distintos (um cão em cada linha) da base de dados *DogFaceNet*.

Fonte: Elaborada pelo autor.

Jia (2019), pois não possuem os mesmos cenários.

Nos experimentos realizados nesta dissertação, foram avaliadas duas arquiteturas, a *ResNet-50* e a *EfficientFormer-L1*, que foram combinadas com a função de erro *ArcFace*. Em relação a função de erro *ArcFace*, foram utilizados 3 subcentros para a realização do aprendizado e os hiperparâmetros foram definidos conforme proposto por Zhang et al. (2019), sendo que o parâmetro de margem é fixo $m = 0.5$ e o parâmetro de escala é obtido dinamicamente através da Equação 4, em que C é o número de classes.

$$s = \sqrt{2} \cdot \log(C - 1) \tag{4}$$

Ainda, conforme proposto por Mougeot, Li e Jia (2019), o protocolo utilizado para validação da arquitetura segue a estratégia *open-set*. Dessa forma, para a utilização do conjunto de testes e sua divisão em *probe* e *gallery*, foi utilizado um parâmetro m, definido em (5), sendo responsável por definir a quantidade de elementos utilizados na *gallery* e, consequentemente,

Figura 32 – Imagens de quatro cães distintos da base de dados *Flickr-dog Dataset*, sendo na linha superior dois indivíduos da raça Pug e na linha inferior dois indivíduos da raça Rusky.

Fonte: Moreira et al. (2017)

Figura 33 – Imagens de dois cães distintos da base de dados *Snoopybook Dataset*.

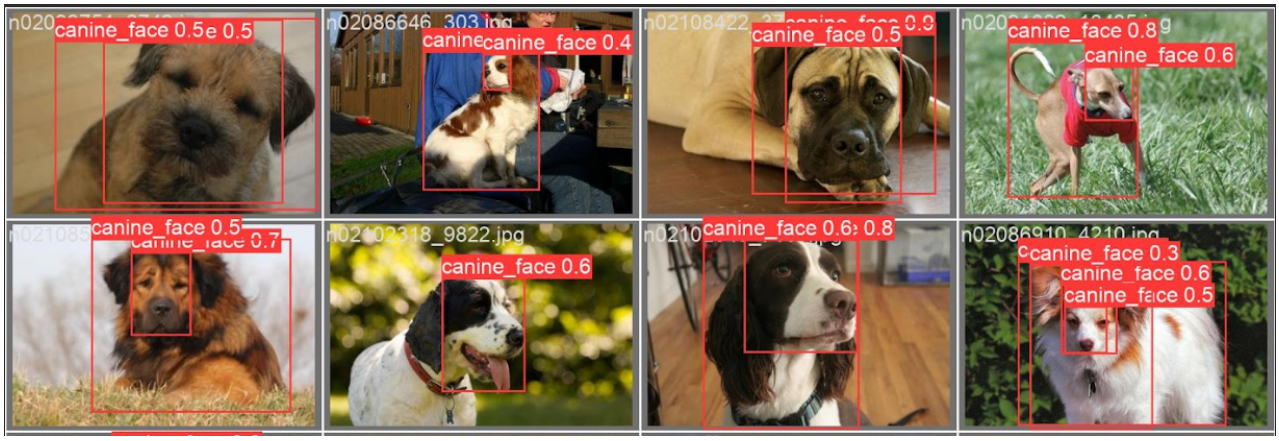
Fonte: Moreira et al. (2017).

os demais elementos são utilizados no conjunto para avaliação.

$$0 < m \leq 5 \tag{5}$$

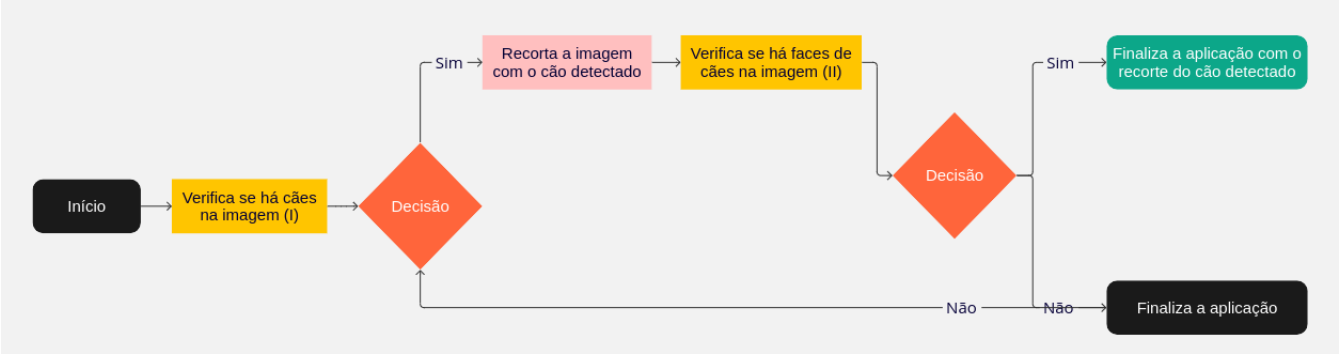
Outrossim, ainda na avaliação para cenários de identificação, também foi utilizada a validação cruzada com k igual a 10. Também, para a verificação de comparações genuínas e impostoras, foram utilizados 2.500 pares positivos e 2.500 pares negativos de imagens de cães,

Figura 34 – Exemplos de resultados da aplicação do *Autodistill* em algumas imagens da base de dados *DogID Dataset*, para a rotularização da base de dados e a detecção das faces dos cães nas imagens.



Fonte: Elaborada pelo autor.

Figura 35 – Arquitetura de detecção de cães e faces de cães em imagens da *DogID Dataset*.



Fonte: Elaborada pelo autor.

assim balanceando as métricas.

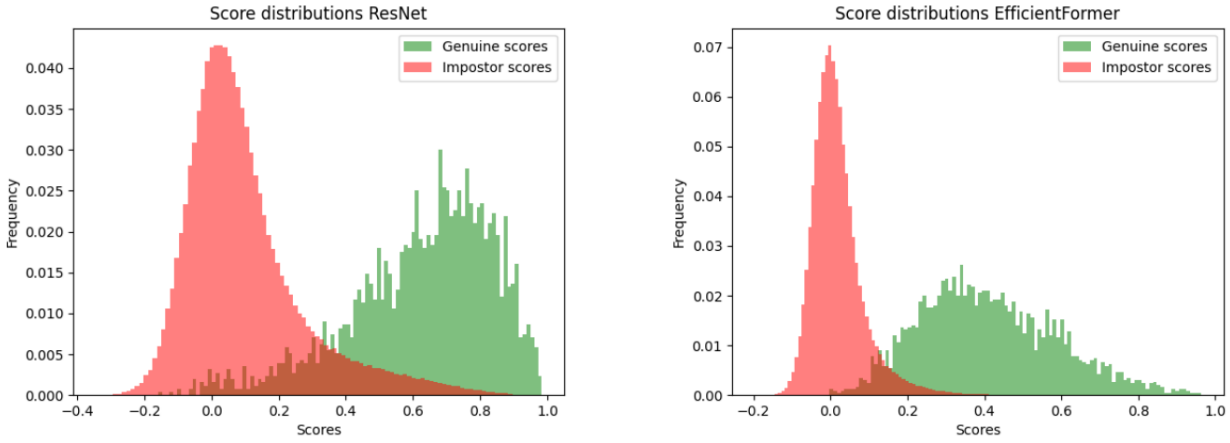
Em relação às métricas, no cenário de verificação, foram utilizadas a *Area Under the Curve* (AUC), curvas ROC, *Equal Error Rate*, a acurácia, a F1-Score, a Precisão e a Revocação. Em relação ao cenário de identificação, há o uso do classificador *Support Vector Machine* (SVM) e das métricas com a curva CMC.

5.4.1.1 Experimentos de Verificação

Para o experimento de verificação foram utilizadas as arquiteturas *ResNet-50* e *EfficientFormer-L1*, com os parâmetros definidos na seção 5.4.1.

A Figura 36 mostra as distribuições dos *scores* genuínos e impostores. Observa-se claramente que utilizando-se a arquitetura *ResNet-50*, as distribuições de *scores* entre genuínos e impostores estão mais próximas, o que gerou uma maior área de intersecção e, consequentemente, maiores taxas de erro. Por outro lado, utilizando-se a arquitetura *EfficientFormer-L1*, as distribuições dos *scores* das comparações genuínas e impostoras estão mais distantes e, com isso, há uma menor área de intersecção gerando menores taxas de erros.

Figura 36 – Distribuições dos *scores* das comparações genuínas e impostoras, utilizando a extração de características por meio das arquiteturas (a) *ResNet-50* e (b) *EfficientFormer-L1*, para a base de dados *DogFaceNet*.



(a) ResNet-50 (b) EfficientFormer-L1

Fonte: Elaborada pelo autor.

A partir das distribuições dos *scores* genuínos e impostores, foram geradas as curvas ROC e calculadas as métricas de AUC, *Equal Error Rate* (EER), acurácia, F1-*score*, precisão e revocação. Os valores destas métricas, para ambas as arquiteturas, *ResNet-50* e *EfficientFormer-L1*, estão apresentadas na Tabela 3.

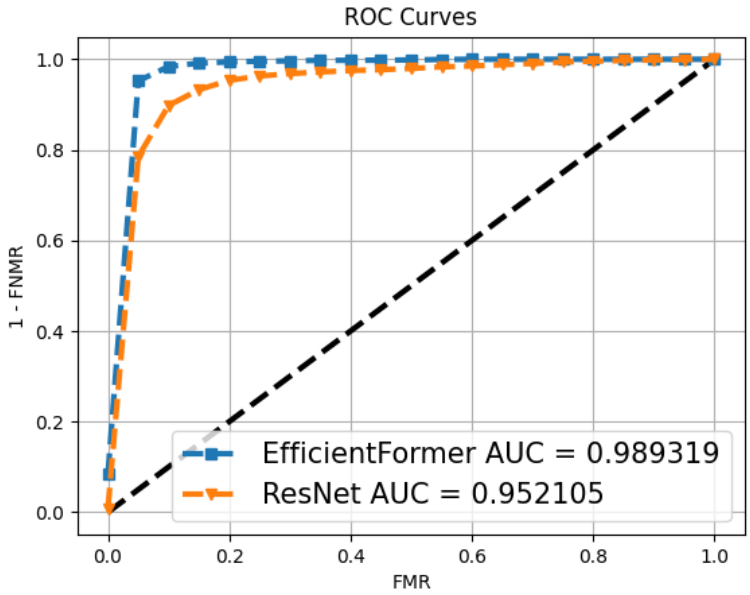
Tabela 3 – Métricas obtidas na tarefa de verificação utilizando a base de dados *DogFaceNet* com dados de características extraídas das arquiteturas *ResNet-50* e *EfficientFormer-L1*

	AUC	EER THR	EER	Accuracy	F1-Score	Precision	Recall
EfficientFormer-L1	0.989319	0.145812	0.048873	0.961147	0.960757	0.970517	0.951190
ResNet-50	0.952105	0.333500	0.100762	0.922100	0.920247	0.942669	0.898867

Conforme pode-se observar na Tabela 3, a arquitetura *EfficientFormer-L1* teve um melhor desempenho sobre a arquitetura *ResNet-50*, apresentando maiores valores de AUC, acurácia, F1-Score, precisão e revocação, e um menor valor de EER. Os valores de AUC, por exemplo, foram 0,952105 e 0,989319 para a *ResNet-50* e *EfficientFormer-L1*, respectivamente.

A Figura 37 mostra as curvas ROC para as arquiteturas utilizadas. Observa-se também por estas curvas que o desempenho da arquitetura *EfficientFormer-L1* foi superior à *ResNet-50*.

Figura 37 – Curvas ROC obtidas na tarefa de verificação pelas arquiteturas *ResNet-50* e *EfficientFormer-L1* utilizando a base de dados *DogFaceNet*.



Fonte: Elaborada pelo autor.

Conforme já mencionado na Seção 3.4, Mougeot, Li e Jia (2019) obtiveram uma taxa de 92% na tarefa de verificação sobre a base de dados *DogFaceNet* reduzida, também utilizando a *ResNet-50*.

Além disso, Yoon, So e Rhee (2021), obtiveram uma taxa média de 88% também na tarefa de verificação sobre a mesma base de dados, utilizando *ArcFace* e o comprimento do vetor, do inglês *Vector Length* (VL).

Embora não seja possível uma comparação direta, esses resultados obtidos pelos trabalhos correlatos contribuem para corroborar o resultado obtido neste trabalho que aponta a superioridade da arquitetura baseada em *Visual Transformer* sobre a arquitetura baseada em *Convolutional Neural Network*.

5.4.1.2 Experimentos de Identificação

Para o experimento de identificação também foram utilizadas as arquiteturas *ResNet-50* e *EfficientFormer-L1*, com os parâmetros definidos na Seção 5.4.1. Para a realização da avaliação, duas diferentes estratégias para divisão dos dados foram abordadas.

Primeiramente, foi calculada a curva *Cumulative Matching Characteristic* (CMC)

utilizando a distância do cosseno entre os elementos do conjunto *probe* com todos os elementos presentes no *gallery*.

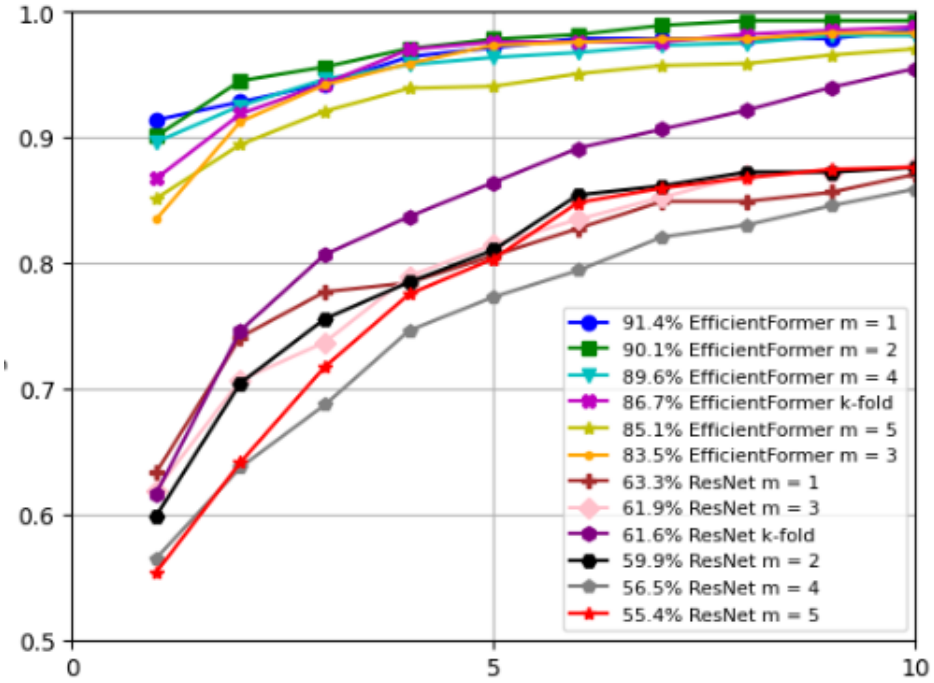
Ainda, a base foi dividida utilizando a estratégia de validação cruzada. Com isso, 90% dos dados do conjunto de dados abertos, proposto por Mougeot, Li e Jia (2019), foram utilizados para a criação do conjunto de *gallery* e 10% para avaliação. Para a classificação, foi utilizado o algoritmo *Support Vector Machines* (SVM). A Tabela 4 mostra os resultados obtidos por ambas as arquiteturas.

Tabela 4 – Resultados das acurácias obtidas pelas arquiteturas *ResNet-50* e *EfficientFormer-L1* na tarefa de identificação seguindo o protocolo de validação cruzada com $k = 10$, na base de dados *DogFaceNet*.

	Protocolo de Validação Cruzada (*K-Fold*)
EfficientFormer-L1	0.8834 ± 0.0356
ResNet-50	0.4653 ± 0.0485

Na complementação dos resultados, a Figura 38 apresenta as curvas CMC para cada protocolo de separação dos dados e para cada uma das arquiteturas avaliadas neste trabalho.

Figura 38 – Curvas CMC para cada protocolo de separação de dados e para cada uma das arquiteturas avaliadas neste trabalho (*ResNet-50* e *EfficientFormer-L1*), na base de dados *DogFaceNet*.



Fonte: Elaborada pelo autor.

A partir das curvas CMC apresentadas na Figura 38, observa-se que a arquitetura *EfficientFormer-L1* apresentou desempenhos muito superiores aos desempenhos da *ResNet-50*, considerando a tarefa de identificação. O maior valor de acurácia *rank-1* foi de 91,4% com a *EfficientFormer-L1*, com m igual a 1, ao passo que a maior acurácia *rank-1* obtida pela arquitetura *ResNet-50* foi de 63,3%, também com m igual a 1.

Conforme mencionado na Seção 3.4, Mougeot, Li e Jia (2019) obtiveram uma acurácia *rank-1* de 60,44% na tarefa de identificação, utilizando a arquitetura *ResNet-50*, sobre a base de dados *DogFaceNet* reduzida.

5.4.2 Experimentos na Base de Dados *DogID Dataset*

Nesta seção são descritos os experimentos realizados na base de dados *DogID Dataset*, proposta neste trabalho e detalhada na Subseção 5.2.1. A base possui 125.873 imagens de 39.148 cães, sendo 90% da base para treino e 10% para validação. Ainda, as imagens utilizadas nos experimentos são resultantes da obtenção por meio do módulo de detecção de faces caninas detalhados na Seção 5.3. A Figura 39 mostra imagens faciais de três cães da base de dados *DogID Dataset* (um cão por linha) obtidas após o processo de detecção e recorte das faces. É possível observar por meio destas amostras que a base de dados apresenta desafios que podem dificultar o reconhecimento facial dos animais, tais como: variações de pose, iluminação e expressões faciais, além de obstruções.

Para a avaliação no cenário de verificação, foram selecionados 20.000 pares positivos e 20.000 pares negativos, sendo os positivos imagens faciais do mesmo cão e os negativos imagens faciais de cães distintos.

Assim como nos experimentos com a base de dados *DogFaceNet*, aqui também foram avaliadas as duas arquiteturas, *ResNet-50* e *EfficientFormer-L1*, que também foram combinadas com a função de erro *ArcFace*. Ainda, os parâmetros utilizados nos experimentos com a base *DogID Dataset*, foram os mesmos utilizados nos experimentos com a base *DogFaceNet*, como, por exemplo, o uso de 3 subcentros no *ArcFace* e o uso dos hiperparâmetros propostos por Zhang et al. (2019).

As métricas nos cenários de verificação e identificação também foram as mesmas utilizadas nos experimentos com a base de dados *DogFaceNet*.

5.4.2.1 Experimentos de Verificação

A Figura 40, mostra as distribuições dos *scores* das comparações genuínas e impostoras. Assim como no experimento realizado na *DogFaceNet*, com a utilização da arquitetura *ResNet-50*, as distribuições de *scores* entre genuínos e impostores gerou uma maior intersecção e, consequentemente, maiores taxas de erro. Já a arquitetura *EfficientFormer-L1* teve uma distri-

Figura 39 – Imagens faciais de 3 cães da base de dados *DogID Dataset* (um animal em cada linha), obtidas após os processos de detecção facial e recorte.

Fonte: Elaborada pelo autor.

buição de *scores* entre genuínos e impostores com uma intersecção menor, consequentemente, gerando menores taxas de erro.

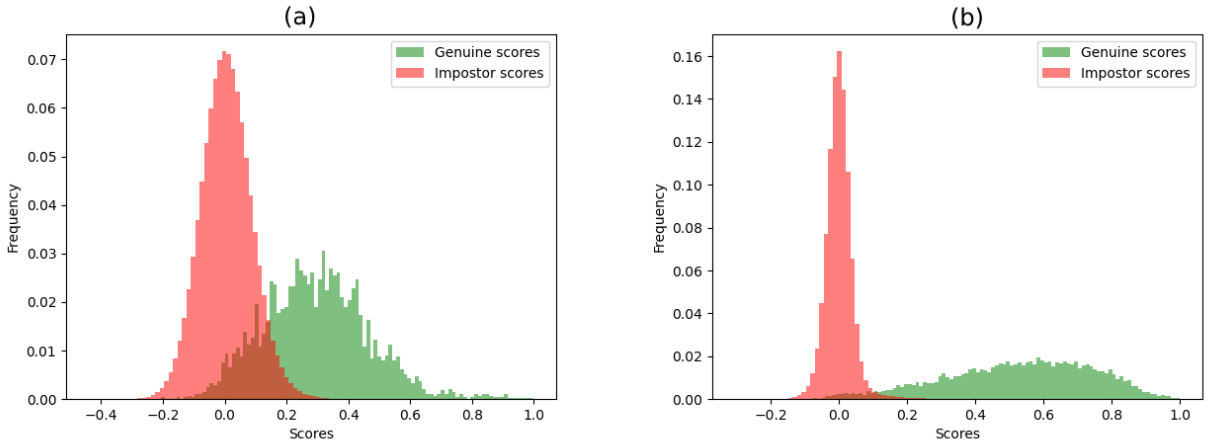
A partir das distribuições dos *scores* genuínos e impostores, foram geradas as curvas ROC e calculadas as métricas de AUC, *Equal Error Rate* (EER), acurácia, F1-*score*, precisão e revocação. Os valores destas métricas, para ambas as arquiteturas, *ResNet-50* e *EfficientFormer-L1*, estão apresentadas na Tabela 5.

Tabela 5 – Métricas obtidas na tarefa de verificação utilizando a base de dados *DogID Dataset* com dados de características extraídas das arquiteturas *ResNet-50* e *EfficientFormer-L1*

	AUC	**EER THR**	**EER**	**Acurácia**	**F1-Score**	**Precisão**	**Recall**
EfficientFormer-L1	0.989803	0.079167	0.026745	0.965418	0.966851	0.988701	0.945946
ResNet-50	0.949456	0.117671	0.117638	0.903846	0.933443	0.972340	0.897435

Conforme pode-se observar na Tabela 5, a arquitetura *EfficientFormer-L1* teve um melhor desempenho sobre a arquitetura *ResNet-50*, apresentando maiores valores de AUC, acurácia, F1-Score, precisão e revocação, e um menor valor de EER. Os valores de AUC, por exemplo, foram 0,949456 e 0,989803 para a *ResNet-50* e *EfficientFormer-L1*, respectivamente.

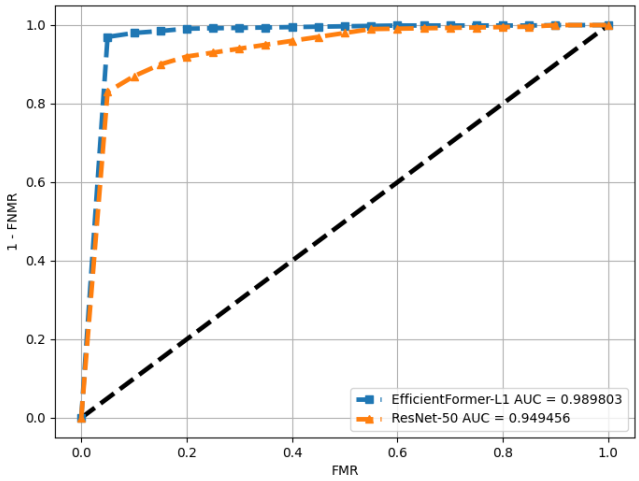
Figura 40 – Distribuições dos *scores* das comparações genuínas e impostoras, utilizando a extração de características por meio das arquiteturas (a) *ResNet-50* e (b) *EfficientFormer-L1*, para a base de dados *DogID Dataset*



Fonte: Elaborada pelo autor.

A Figura 41 mostra as curvas ROC para as arquiteturas utilizadas. Observa-se também por estas curvas que o desempenho da arquitetura *EfficientFormer-L1* foi superior à *ResNet-50*.

Figura 41 – Curvas ROC obtidas na tarefa de verificação pelas arquiteturas *ResNet-50* e *EfficientFormer-L1* utilizando a base de dados *DogID Dataset*



Fonte: Elaborada pelo autor.

5.4.2.2 Experimentos de Identificação

Para o experimento de identificação também foram utilizadas as arquiteturas *ResNet-50* e *EfficientFormer-L1*, com os parâmetros definidos na Seção 5.4.2.

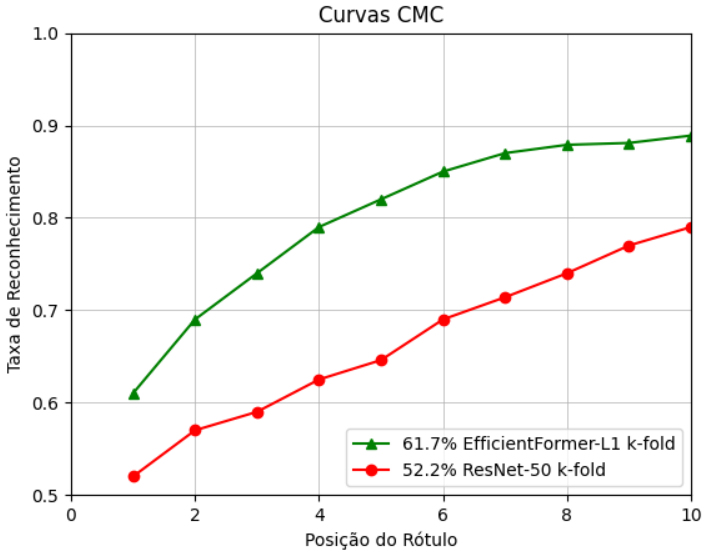
Para a divisão dos dados foi utilizada a validação cruzada, com $k = 10$. Quanto ao algoritmo de classificação, foi utilizado o *Support Vector Machines* (SVM). A Tabela 6 mostra os resultados obtidos utilizando-se o protocolo de validação cruzada, com o classificador SVM, para ambas as arquiteturas.

Tabela 6 – Resultados das acurácias obtidas pelas arquiteturas *ResNet-50* e *EfficientFormer-L1* na tarefa de identificação seguindo o protocolo de validação cruzada com $k = 10$, na base de dados *DogID Dataset*.

	Protocolo de Validação Cruzada (*K-Fold*)
EfficientFormer-L1	0.6822 ± 0.1222
ResNet-50	0.5259 ± 0.1678

Na complementação dos resultados, a Figura 42 apresenta as curvas CMC ambas as arquiteturas avaliadas neste trabalho.

Figura 42 – Curvas CMC obtidas pelas arquiteturas *ResNet-50* e *EfficientFormer-L1* na base de dados *DogID Dataset*.



Fonte: Elaborada pelo autor.

A partir das curvas CMC apresentadas na Figura 42, observa-se que a arquitetura *EfficientFormer-L1* obteve também na base de dados *DogID Dataset* um desempenho superior ao da *ResNet-50*, considerando a tarefa de identificação.

6 Conclusão

Com o avanço da tecnologia, principalmente no cenário de cidades inteligentes e dispositivos de Internet das Coisas, há uma melhoria e, consequentemente, sofisticação em aplicações de biometria. Com isso, é possível explorar o campo de biometria animal, pois ainda há poucas aplicações e estudos na área aplicados a cães, conforme o estudo realizado neste trabalho.

Conforme apontado neste trabalho, o número de animais domésticos está crescendo expressivamente e, consequentemente, o seu mercado de atendimentos e produtos, como por exemplo, em estabelecimentos veterinários e planos de saúde. Dessa forma, é essencial uma forma de rastrear e realizar o controle desses animais, dentre eles os cães.

Com o uso de biometria é possível ter rastreabilidade e gestão no controle de cães. Isso traz avanços significativos tanto para localização, verificação da saúde e, também na diminuição de fraudes.

Quanto às técnicas de reconhecimento, os estudos e experimentos realizados indicam que com o uso de aprendizagem de máquina profunda é possível desenvolver tecnologias eficazes e eficientes para realizar o reconhecimento biométrico de animais, particularmente cães, por meio de suas faces, proporcionando, assim, formas de monitorar esses animais em cidades, de encontrar animais perdidos, realizar controle de doenças de forma mais efetiva e auxiliar na prevenção ou detecção de fraudes.

Além disso, nota-se que o uso e o avanço de arquiteturas de Aprendizado de Máquina Profundo, como por exemplo, *Vision Transformers* tem proporcionado resultados significativos no campo de Visão Computacional. Por meio dos experimentos apresentados neste trabalho, o uso de ViT superou o uso de arquiteturas consideradas até o momento estado da arte de Visão Computacional, no caso as Redes Neurais Convolucionais, mais especificamente a arquitetura *ResNet*.

6.1 Contribuições deste Trabalho

Com este trabalho, foi possível ter, em resumo, as seguintes contribuições:

- Proposta de um método robusto e eficaz, baseado em técnicas do estado da arte de Aprendizado de Máquina Profundo para a identificação biométrica de animais, especificamente em cães, por meio de suas características faciais;
- Criação de uma base de dados de cães, a *DogID Dataset*, que será disponibilizada publicamente para outros pesquisadores e, assim, proporcionar avanços em estudos da

área. Além disso, a base possui 125.873 imagens de 39.148 cães, sendo pelo menos duas imagens por animal. Portanto, sendo, a maior base de dados disponibilizada publicamente com foco em biometria, segundo a revisão bibliográfica realizada neste estudo.

6.2 Direcionamentos para Trabalhos Futuros

Para trabalhos futuros, alguns itens podem ser explorados:

- Utilização dos métodos propostos neste trabalho, baseado em Aprendizado de Máquina Profundo, para a identificação biométrica de outros animais, como por exemplo, gatos;
- Utilização de *Soft Biometrics*, como, por exemplo, a raça, para a identificação biométrica em cães e gatos;
- Com o avanço dos sensores de captura, explorar o uso das características biométricas dos focinhos para a identificação animal, em particular dos gatos e cachorros.

6.3 Publicação Realizada

Resultados parciais desta dissertação de mestrado foram descritos em um artigo científico (CANTO et al., 2023) aceito para apresentação e publicação na 12th Brazilian Conference on Intelligent Systems (BRACIS 2023).

Referências

ALLEN, A.; GOLDEN, B.; TAYLOR, M.; PATTERSON, D.; HENRIKSEN, D.; SKUCE, R. Evaluation of retinal imaging technology for the biometric identification of bovine animals in northern ireland. *Livestock science*, Elsevier, v. 116, n. 1-3, p. 42–52, 2008.

CANTO, V. H. B.; MANESCO, J. R. R.; SOUZA, G. B.; MARANA, A. N. *Lecture Notes in Computer Science*. Belo Horizonte, Minas Gerais, Brazil: Springer, 2023. Brazilian Conference on Intelligent Systems (BRACIS). ISBN 978-3-031-45388-5.

CHOLLET, F. How convolutional neural networks see the world. *The Keras Blog*, v. 30, 2016.

DAUGMAN, J. How iris recognition works. In: *The essential guide to image processing*. [S.l.]: Elsevier, 2009. p. 715–739.

DENG, J.; GUO, J.; LIU, T.; GONG, M.; ZAFEIRIOU, S. Sub-center arcface: Boosting face recognition by large-scale noisy web faces. In: SPRINGER. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*. [S.l.], 2020. p. 741–757.

DENG, J.; GUO, J.; XUE, N.; ZAFEIRIOU, S. Arcface: Additive angular margin loss for deep face recognition. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2019. p. 4690–4699.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

GEEKSFORGEEKS. *Residual Networks (ResNet) - Deep Learning*. 2023. <https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning>. Acesso em 19 de Junho de 2023.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016.

GOOGLEDEVELOPERS. *Google Developers*. 2023. <https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks>. Acesso em 14 de Junho de 2023.

GROUNDED-SAM. *Grounded-SAM*. 2023. <https://github.com/IDEA-Research/Grounded-Segment-Anything>. Acesso em 21 de Junho de 2023.

GU, J.; WANG, Z.; KUEN, J.; MA, L.; SHAHROUDY, A.; SHUAI, B.; LIU, T.; WANG, X.; WANG, G.; CAI, J. et al. Recent advances in convolutional neural networks. *Pattern recognition*, Elsevier, v. 77, p. 354–377, 2018.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Identity mappings in deep residual networks. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 630–645.

IPB. *Censo Pet IPB: com alta recorde de 6% em um ano, gatos lideram crescimento de animais de estimação no Brasil*. 2021. <https://institutopetbrasil.com/fique-por-dentro/amor-pelos-animais-impulsiona-os-negocios-2-2/>. Acesso em 06 de Junho de 2023.

IPB. *Varejo Pet em crescimento*. 2022. <https://caesegatos.com.br/marco-historico-varejo-pet-cresce-164-em-2022-e-fatura-r-602-bilhoes>. Acesso em 04 de Junho de 2023.

JAIN, A. K.; LI, S. Z. *Handbook of face recognition*. [S.l.]: Springer, 2011. v. 1.

JANG, D.-H.; KWON, K.-S.; KIM, J.-K.; YANG, K.-Y.; KIM, J.-B. Dog identification method based on muzzle pattern image. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 10, n. 24, p. 8994, 2020.

KIM, P. Convolutional neural network. In: *MATLAB deep learning*. [S.l.]: Springer, 2017. p. 121–147.

KORTLI, Y.; JRIDI, M.; FALOU, A. A.; ATRI, M. Face recognition systems: A survey. *Sensors*, MDPI, v. 20, n. 2, p. 342, 2020.

KUMAR, S.; SINGH, S. K.; SINGH, R.; SINGH, A. K. *Animal Biometrics: Techniques and Applications*. [S.l.]: Springer, 2018.

LAI, K.; TU, X.; YANUSHKEVICH, S. Dog identification using soft biometrics and neural networks. In: IEEE. *2019 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2019. p. 1–8.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 1998.

LI, S.; JIAO, J.; HAN, Y.; WEISSMAN, T. Demystifying resnet. *arXiv preprint arXiv:1611.01186*, 2016.

LI, Y.; YUAN, G.; WEN, Y.; HU, J.; EVANGELIDIS, G.; TULYAKOV, S.; WANG, Y.; REN, J. Efficientformer: Vision transformers at mobilenet speed. *Advances in Neural Information Processing Systems*, v. 35, p. 12934–12949, 2022.

LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, IEEE, 2021.

LIU, S.; ZENG, Z.; REN, T.; LI, F.; ZHANG, H.; YANG, J.; LI, C.; YANG, J.; SU, H.; ZHU, J. et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. *arXiv preprint arXiv:2303.05499*, 2023.

MOREIRA, T. P.; PEREZ, M. L.; WERNECK, R. de O.; VALLE, E. Where is my puppy? retrieving lost dogs by facial features. *Multimedia Tools and Applications*, Springer, v. 76, n. 14, p. 15325–15340, 2017.

MOUGEOT, G.; LI, D.; JIA, S. A deep learning approach for dog face verification and recognition. In: SPRINGER. *Pacific Rim International Conference on Artificial Intelligence*. [S.l.], 2019. p. 418–430.

RATHGEB, C.; PÖPPELMANN, K.; GONZALEZ-SOSA, E. Biometric technologies for elearning: State-of-the-art, issues and challenges. In: IEEE. *2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA)*. [S.l.], 2020. p. 558–563.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 779–788.

TARG, S.; ALMEIDA, D.; LYMAN, K. Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*, 2016.

TERRA. *Poodle anão é menor que o toy? Conheça os tipos da raça*. [S.l.]: Terra, 2014. <https://www.terra.com.br/vida-e-estilo/mulher/comportamento/poodle-anao-e-menor-que-o-toy-conheca-os-tipos-da-raca,51e4546449787410VgnVCM4000009bcceb0aRCRD.html>. Acesso em 11 de Junho de 2023.

TERVEN, J.; CORDOVA-ESPARZA, D. A comprehensive review of yolo: From yolov1 to yolov8 and beyond. *arXiv preprint arXiv:2304.00501*, 2023.

TROKIELEWICZ, M.; SZADKOWSKI, M. Iris and periocular recognition in arabian race horses using deep convolutional neural networks. In: IEEE. *2017 IEEE International Joint Conference on Biometrics (IJCB)*. [S.l.], 2017. p. 510–516.

TU, X.; LAI, K.; YANUSHKEVICH, S. Transfer learning on convolutional neural networks for dog identification. In: IEEE. *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*. [S.l.], 2018. p. 357–360.

ULTRALYTICS. *Ultralytics*. 2023. <https://docs.autodistill.com/>. Acesso em 23 de Junho de 2023.

USP, J. *Cresce o número de adoções e de abandono de animais na pandemia*. 2021. <https://jornal.usp.br/atualidades/cresce-o-numero-de-adocoes-e-de-abandono-de-animais-na-pandemia/>. Acesso em 24 de Maio de 2023.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017.

YOON, B.; SO, H.; RHEE, J. A methodology for utilizing vector space to improve the performance of a dog face identification model. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 11, n. 5, p. 2074, 2021.

YOUTALK-INSURANCE. *Pet insurance fraud increases*. 2018. <https://youtalk-insurance.com/broker-news/400-rise-in-pet-insurance-fraud-highlights-need-for-new-approach>. Acesso em 27 de Maio de 2023.

ZHANG, D. D. *Automated biometrics: Technologies and systems*. [S.l.]: Springer Science & Business Media, 2013. v. 7.

ZHANG, K.; SUN, M.; HAN, T. X.; YUAN, X.; GUO, L.; LIU, T. Residual networks of residual networks: Multilevel residual networks. *IEEE Transactions on Circuits and Systems for Video Technology*, IEEE, v. 28, n. 6, p. 1303–1314, 2017.

ZHANG, X.; ZHAO, R.; QIAO, Y.; WANG, X.; LI, H. Adacos: Adaptively scaling cosine logits for effectively learning deep face representations. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2019. p. 10823–10832.