



**UNIVERSIDADE FEDERAL DO MARANHÃO**

**BACHARELADO INTERDISCIPLINAR EM CIÊNCIA E TECNOLOGIA**

**COORDENAÇÃO DO CURSO DE ENGENHARIA DA COMPUTAÇÃO - CCEC**

**HAROLDO GOMES BARROSO FILHO**

**PROCESSAMENTO DE IMAGENS**

**RUAN PACTRICK DE SOUSA E SOUSA (2022013733)**

**APLICAÇÃO DE IDENTIFICAR A RAÇA DO CACHORRO E GATO**

**SÃO LUÍS (MA)**

**2026**

## 1. SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>3</b>
<b>2. REFERÊNCIAL TEORICO .....</b>	<b>4</b>
2.0.1    INTERSEÇÃO ENTRE PDI E VISÃO COMPUTACIONAL.....	4
2.0.2    A IMAGEM COMO MATRIZ DE CARACTERÍSTICAS .....	5
2.0.3    FILTRAGEM ESPACIAL E CONVOLUÇÃO.....	5
2.0.4    ARQUITETURA MOBILENETV2.....	6
<b>3. METODOLOGIA .....</b>	<b>7</b>
3.0.1    CONJUNTO DE DADOS (DATASET) .....	7
3.0.2    PIPELINE DE PROCESSAMENTO DA IMAGEM.....	7
3.0.3    TREINAMENTO E TRANSFER LEARNING .....	7
<b>4. RESULTADOS E ANÁLISE .....</b>	<b>8</b>
4.0.1    DISCRIMINAÇÃO DE RAÇAS SIMILARES.....	8
4.0.2    INFLUÊNCIA DA RESOLUÇÃO E FOCO .....	9
<b>5. CONCLUSÃO .....</b>	<b>10</b>
<b>6. REFERÊNCIAS .....</b>	<b>11</b>

## 1. INTRODUÇÃO

A inteligência artificial, em particular na subárea de Visão Computacional, tem alcançado resultados expressivos na tarefa de classificação de imagens, possibilitando avanços significativos em diversas áreas do conhecimento. No contexto doméstico e veterinário, a identificação correta da raça de um animal assume papel fundamental, uma vez que diferentes raças apresentam predisposições genéticas a determinadas patologias, demandas nutricionais específicas e variações comportamentais relevantes.

Entretanto, para tutores não especialistas, a distinção entre raças com fenótipos semelhantes, como o American Pit Bull Terrier e o Staffordshire Bull Terrier, pode representar um desafio considerável. Essa dificuldade evidencia a necessidade de soluções automatizadas capazes de apoiar a identificação precisa a partir de características visuais sutis.

Diante desse cenário, o objetivo deste trabalho consiste no desenvolvimento de um sistema automatizado, acessível e computacionalmente eficiente, capaz de auxiliar na identificação de raças de cães e gatos a partir de uma única fotografia. Para viabilizar sua execução em dispositivos com recursos de processamento limitados, como smartphones, optou-se pelo emprego de arquiteturas de redes neurais convolucionais otimizadas.

A justificativa para este desenvolvimento fundamenta-se na democratização do acesso a ferramentas de diagnóstico visual baseadas em inteligência artificial, permitindo que estudantes, profissionais da área veterinária e tutores obtenham informações rápidas e confiáveis sobre o perfil racial dos animais. Dessa forma, o sistema proposto contribui para a tomada de decisões mais informadas relacionadas aos cuidados, à saúde e ao bem-estar animal.

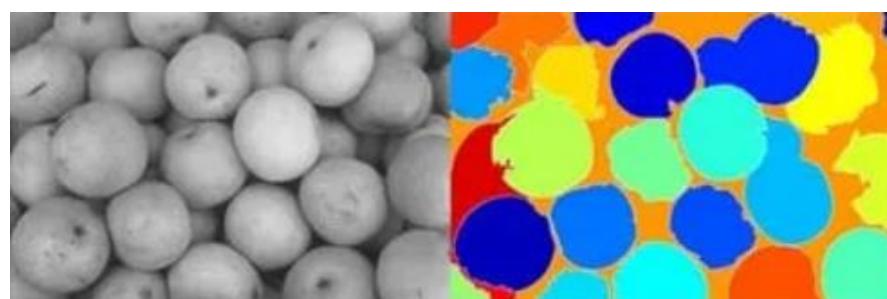
## 2. REFERÊNCIAL TEORICO

A fundamentação teórica deste estudo combina conceitos tradicionais de Processamento Digital de Imagens (PDI) com métodos modernos de Aprendizado Profundo (Deep Learning). Para que um sistema computacional consiga diferenciar classes visuais complexas, como raças de animais, é preciso entender primeiro como a informação visual é discretizada, representada de forma matemática e processada. Neste tópico, discutiremos a natureza matricial da imagem digital, as operações de filtragem espacial que possibilitam a extração de características (convolução) e a arquitetura de rede neural otimizada empregada na classificação.

### 2.0.1 INTERSEÇÃO ENTRE PDI E VISÃO COMPUTACIONAL

O Processamento Digital de Imagens (PDI) e a Visão Computacional são campos diferentes, mas altamente conectados, com influência recíproca no avanço de soluções baseadas em imagens. O PDI se concentra principalmente na manipulação e transformação de imagens, ao passo que a Visão Computacional emprega essas técnicas para extrair informações semânticas, com o objetivo de interpretar e entender a cena representada.

Figura 01 – Exemplo de Processamento de Imagens.



Fonte: [SlideShare de Processamento de Imagens](#)

Figura 02 – Exemplo de Visão Computacional



Fonte: [Blog da Geo](#)

Neste projeto, o PDI é essencial, pois oferece a base necessária para o funcionamento eficiente dos algoritmos de reconhecimento de padrões, principalmente por meio de processos como a normalização e a reamostragem das imagens. Essas etapas são fundamentais para minimizar variações indesejadas e assegurar maior consistência nos dados de entrada.

O problema tratado é classificado na literatura como Classificação Visual de Granulação Fina (Fine-Grained Visual Categorization – FGVC), uma tarefa complexa que envolve a distinção de classes subordinadas que são visualmente semelhantes. Esse tipo de classificação requer sensibilidade a variações delicadas em forma, textura e padrões visuais, o que torna essencial a combinação eficaz entre técnicas de PDI e métodos avançados de Visão Computacional.

## 2.0.2 A IMAGEM COMO MATRIZ DE CARACTERÍSTICAS

Do ponto de vista computacional, uma imagem digital  $I$  pode ser representada como uma função bidimensional discretizada  $f(x,y)$ , na qual o valor da função em cada par de coordenadas  $(x,y)$  corresponde à intensidade luminosa, ou brilho, daquele ponto da imagem.

No contexto da identificação de raças, a informação espectral, especialmente a cor, desempenha um papel fundamental. Para esse propósito, adota-se o espaço de cores RGB, no qual cada pixel é representado não por um valor escalar, mas por um vetor tridimensional composto pelas componentes vermelho (R), verde (G) e azul (B). Dessa forma, a imagem é modelada como um tensor de dimensões:

$$\text{Fórmula: Altura} \times \text{Largura} \times 3.$$

Além da informação cromática, a textura da pelagem do animal constitui uma característica visual essencial para a diferenciação entre raças. No domínio espacial, a textura pode ser interpretada como a taxa de variação da intensidade entre pixels vizinhos. Regiões que apresentam variações abruptas estão associadas a altas frequências espaciais, correspondendo a padrões mais detalhados, enquanto variações suaves indicam baixas frequências, relacionadas a regiões mais homogêneas da imagem.

## 2.0.3 FILTRAGEM ESPACIAL E CONVOLUÇÃO

A extração de características visuais fundamenta-se na operação linear de convolução. No contexto do Processamento Digital de Imagens, a convolução é definida pela aplicação de uma máscara, ou *kernel*,  $w$ , de dimensões  $m \times n$  sobre a imagem  $f$ , resultando em uma nova imagem  $g$ , conforme a expressão:

$$g(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x - s, y - t)$$

Em abordagens clássicas de PDI, filtros como Sobel, utilizados para detecção de bordas, ou filtros de média, empregados para suavização, possuem coeficientes fixos,

definidos analiticamente a partir de princípios matemáticos. Esses filtros são projetados para realçar propriedades específicas da imagem, porém apresentam capacidade limitada de adaptação a diferentes padrões visuais.

Em contraste, nas Redes Neurais Convolucionais (Convolutional Neural Networks – CNNs), os coeficientes da máscara  $w$  são tratados como parâmetros ajustáveis, aprendidos de forma automática e iterativa durante o processo de treinamento. Essa característica permite que o modelo aprenda filtros progressivamente mais complexos, capazes de capturar padrões visuais discriminativos, como o formato específico de uma orelha ou padrões característicos da pelagem, os quais seriam difíceis de serem descritos de forma analítica por meio de filtros tradicionais.

#### **2.0.4 ARQUITETURA MOBILENETV2**

Para viabilizar o processamento da matriz de imagem em dispositivos com restrições de hardware, adotou-se a arquitetura MobileNetV2. Essa rede neural convolucional introduz o conceito de Convoluçãoções Separáveis em Profundidade (Depthwise Separable Convolutions), projetadas para reduzir significativamente o custo computacional sem comprometer de forma substancial o desempenho do modelo.

Diferentemente da convolução padrão, que opera simultaneamente sobre todos os canais de entrada, a convolução separável em profundidade decompõe o processo em duas etapas distintas. Na primeira etapa, denominada filtragem espacial (depthwise convolution), aplica-se um filtro independente a cada canal de entrada, como os canais R, G e B. Em seguida, na etapa de combinação linear (pointwise convolution), utiliza-se uma convolução  $1 \times 1$  para combinar linearmente as saídas obtidas na etapa anterior, gerando novos mapas de características.

Essa decomposição reduz drasticamente a complexidade computacional e o número de operações de ponto flutuante necessárias, tornando o processo de inferência mais rápido e eficiente. Dessa forma, a MobileNetV2 mostra-se especialmente adequada para aplicações em ambientes com recursos limitados, como sistemas web e dispositivos móveis.

### 3. METODOLOGIA

O sistema foi desenvolvido em Python, empregando a biblioteca OpenCV para as etapas de pré-processamento das imagens e o framework TensorFlow para a realização da inferência por meio de redes neurais convolucionais.

#### 3.0.1 CONJUNTO DE DADOS (DATASET)

Foi utilizado o Oxford-IIIT Pet Dataset, contendo imagens rotuladas de 37 raças. Este dataset é ideal para PDI pois apresenta desafios como oclusão (o animal parcialmente escondido), variações de pose e diferentes condições de iluminação.

#### 3.0.2 PIPELINE DE PROCESSAMENTO DA IMAGEM

Antes da classificação, a imagem bruta passa por tratamentos rigorosos:

- **Padronização do Espaço de Cores:** Realiza-se a conversão forçada das imagens para o espaço de cores RGB. Imagens originalmente monocromáticas, em escala de cinza, têm seus valores replicados nos três canais, de modo a garantir a consistência dimensional da entrada do modelo.
- **Reamostragem Geométrica (Redimensionamento):** As imagens são redimensionadas para a resolução de  $160 \times 160$  pixels, utilizando interpolação bilinear. Embora esse procedimento implique na perda de informações de alta frequência, como detalhes muito finos da pelagem, essa resolução apresentou-se como um compromisso adequado entre eficiência computacional e preservação das características morfológicas relevantes para a identificação da raça.
- **Normalização de Intensidade:** Os valores de intensidade originalmente no intervalo  $[0,255]$  são mapeados linearmente para o intervalo  $[-1,1]$ . Esse processo promove a centralização dos dados, facilitando a aprendizagem dos filtros convolucionais e contribuindo para uma melhor detecção de gradientes e bordas durante a inferência.

#### 3.0.3 TREINAMENTO E TRANSFER LEARNING

Empregou-se a técnica de Transfer Learning, na qual uma rede neural previamente treinada para a extração de características visuais gerais é adaptada para uma nova tarefa específica. Nesse contexto, o modelo passou por um processo de Fine-Tuning (ajuste fino), no qual os parâmetros das camadas finais foram refinados de forma iterativa durante o treinamento.

Esse procedimento permitiu que os filtros convolucionais se tornassem progressivamente mais sensíveis às características discriminativas das 37 raças consideradas no estudo. Como resultado, a rede aprendeu a capturar variações sutis de textura, forma e padrões visuais, possibilitando, por exemplo, a diferenciação entre a textura da pelagem de um Golden Retriever e a de um Labrador, apesar de sua semelhança visual.

## 4. RESULTADOS E ANÁLISE

A avaliação do sistema foi realizada de forma qualitativa por meio de uma interface web desenvolvida com o framework Streamlit.

### 4.0.1 DISCRIMINAÇÃO DE RAÇAS SIMILARES

Um dos principais desafios no Processamento Digital de Imagens aplicado à classificação visual é a elevada similaridade interclasse. No sistema desenvolvido, observou-se eficácia na distinção entre raças visualmente semelhantes, como o Beagle e o Basset Hound. Apesar de ambas apresentarem coloração tricolor semelhante, resultando em histogramas de cores próximos, o modelo foi capaz de explorar informações espaciais para realizar a diferenciação correta.

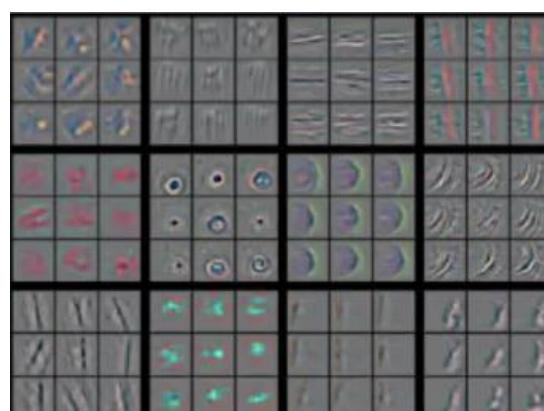
Figura 03 – Foto do Beagle



Fonte: [PetLove](#)

Os filtros convolucionais aprenderam a capturar diferenças morfológicas relevantes, como a proporção e o comprimento das orelhas, bem como o formato geral do corpo. Essas características estruturais, aliadas às informações cromáticas, permitiram ao sistema classificar corretamente as imagens, mesmo em cenários de alta similaridade visual entre classes.

Figura 04 – Exemplo de Filtro Convolucionais



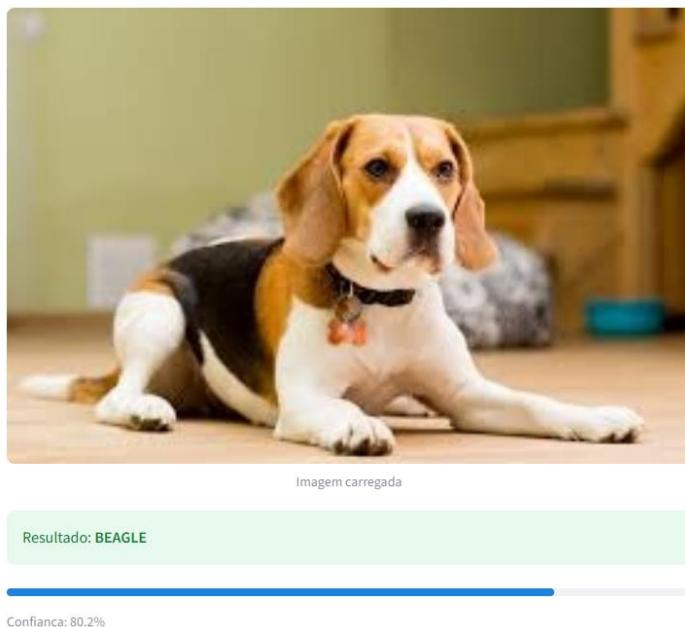
Fonte: [Blog Ateliware](#)

#### 4.0.2 INFLUÊNCIA DA RESOLUÇÃO E FOCO

Observou-se que imagens fora de foco, caracterizadas por desfoque, impactam negativamente o desempenho da classificação em raças cuja diferenciação depende fortemente de informações de textura, como os Terriers de pelo duro. Do ponto de vista do Processamento Digital de Imagens, o desfoque pode ser interpretado como a aplicação de um filtro passa-baixas, que atenua as componentes de alta frequência associadas aos padrões texturais da pelagem.

Como consequência, ocorre a suavização das variações locais de intensidade, resultando na perda de informações discriminativas essenciais para a atuação dos filtros convolucionais. Esse efeito compromete a capacidade do modelo em identificar características específicas da raça, refletindo-se em uma redução do grau de confiança associado às previsões realizadas.

Figura 05 – Teste de confiança no PetID.



Fonte: Própria.

## 5. CONCLUSÃO

O presente trabalho evidenciou a importância das técnicas de Processamento Digital de Imagens como base para a solução de problemas de classificação visual de granulação fina. A manipulação adequada da matriz de imagem, abrangendo etapas como conversão de espaço de cores, reamostragem geométrica e normalização, mostrou-se fundamental para viabilizar a extração eficaz de características discriminativas pelos filtros convolucionais da arquitetura MobileNetV2.

O sistema desenvolvido demonstrou ser capaz de identificar 37 raças de cães e gatos com elevado nível de precisão, corroborando o potencial do Deep Learning como uma ferramenta robusta para análise automática de imagens. Esses resultados validam a aplicabilidade da abordagem proposta em contextos veterinários e biológicos, especialmente em cenários que demandam a distinção entre classes visualmente semelhantes.

## 6. REFERÊNCIAS

- [1] GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento Digital de Imagens**. 3. ed. São Paulo: Pearson, 2010.
- [2] SANDLER, Mark et al. **MobileNetV2: Inverted Residuals and Linear Bottlenecks**. In: CVPR, 2018.
- [3] PARKHI, Omkar M. et al. **Cats and Dogs**. In: IEEE Conference on Computer Vision and Pattern Recognition, 2012.
- [4] LECUN, Y.; BENGIO, Y. **Convolutional networks for images, speech, and time series**. The handbook of brain theory and neural networks, 1995.