Relatório de Pré-processamento de Dados – Dataset RAF-DB

Projeto: Análise Comparativa de Arquiteturas de Deep Learning para Reconhecimento de Expressões Faciais

Data: 14 de setembro de 2025

Autor(a): Luana Tiemann Halicki Cordeiro

Dataset Foco: Real-world Affective Face Database (RAF-DB)

---

1. Introdução e Objetivo

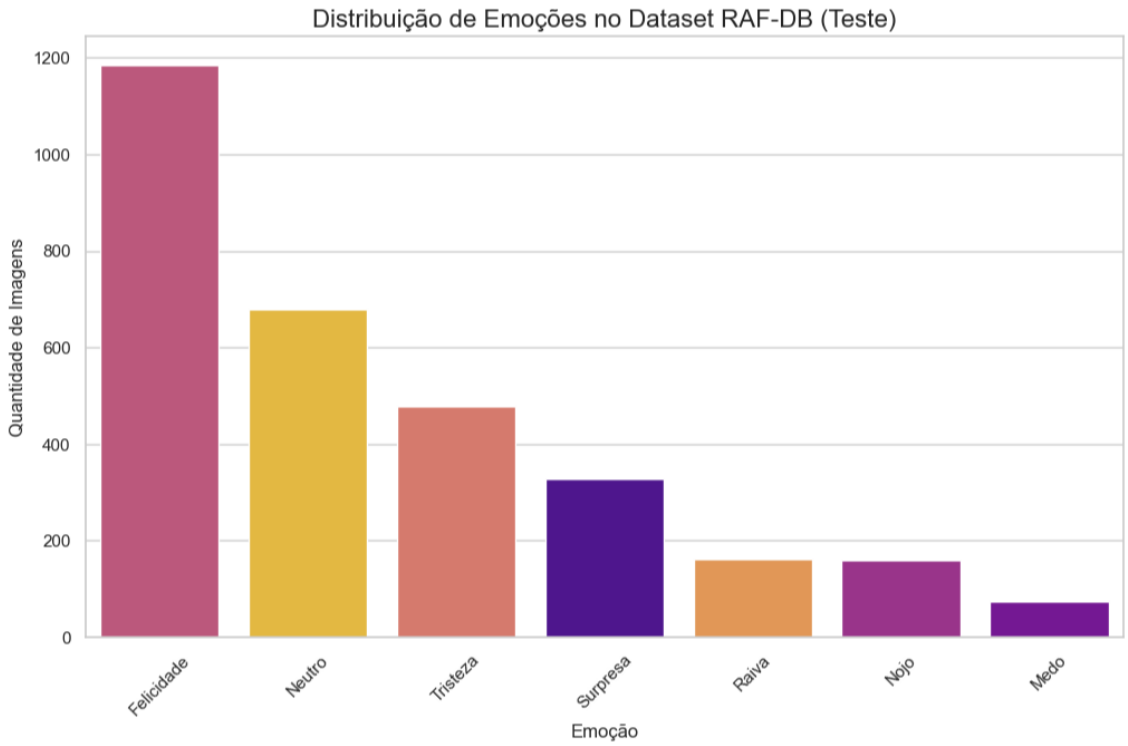
O objetivo desta etapa do projeto foi executar um pipeline metodológico de pré-processamento sobre o dataset RAF-DB. A finalidade foi transformar os dados brutos em um conjunto de imagens otimizado, padronizado e balanceado, adequado para o subsequente treinamento e avaliação de modelos de Deep Learning (ResNet50, EfficientNet, EfficientViT). O desenvolvimento deste pipeline foi informado pelas melhores práticas da literatura e por uma Análise Exploratória de Dados (EDA) preliminar, que identificou as características intrínsecas do dataset e guiou as decisões técnicas subsequentes.

2. Análise Exploratória de Dados (EDA) e Principais Descobertas

Previamente a qualquer transformação, foi conduzida uma análise exploratória para investigar a estrutura, composição e qualidade do dataset RAF-DB. O processo envolveu a catalogação de todos os ficheiros de imagem e seus respectivos rótulos em uma estrutura de dados tabular (DataFrame), permitindo análises quantitativas e qualitativas. A composição geral do dataset foi estabelecida em **15.339 imagens**, divididas em **12.271 para o conjunto de treino** e **3.068 para o de teste**.

Principais Descobertas:

- **Consistência Dimensional:** A análise confirmou que 100% das imagens do dataset possuíam uma resolução padronizada de **100x100 pixels**, validando a alta consistência e qualidade da versão dos dados utilizada.
- **Desbalanceamento de Classes Severo:** Foi identificado um forte desequilíbrio na distribuição das classes no conjunto de treino. A emoção **'Felicidade' demonstrou ser a classe largamente majoritária, com 4.779 instâncias**, enquanto emoções como **'Nojo' (689), 'Raiva' (698) e 'Medo' (272)** se mostraram significativamente sub-representadas.



- **Alta Diversidade e Ambiguidade:** A inspeção visual de amostras revelou uma notável diversidade de sujeitos em termos de idade, etnia e gênero. Contudo, também evidenciou uma considerável ambiguidade visual e sobreposição de características entre certas classes de emoções, notadamente entre os pares 'Medo'/'Surpresa' e 'Raiva'/'Nojo', antecipando os principais desafios para o classificador.

Conclusão da EDA: A análise confirmou a RAF-DB como um dataset de alta qualidade, porém com um viés de distribuição que necessitava de mitigação. A descoberta do desbalanceamento severo foi o principal fator que justificou a adoção de uma estratégia de balanceamento offline.

3. Fase 1: Pré-processamento Pesado (Preparação Offline)

Esta fase foi implementada como um script automatizado, executado uma única vez para gerar um novo dataset processado e estruturado.

Técnicas Aplicadas:

1. Deteção e Alinhamento Facial: Da Imagem Bruta à Pose Normalizada

O passo inicial do pré-processamento consistiu na identificação e normalização da pose facial em cada imagem. Para esta tarefa, foi empregado um detetor facial robusto, o **Multi-task Cascaded Convolutional Networks (MTCNN)**. A escolha

desta técnica foi estratégica, pois, diferentemente de detetores mais simples que retornam apenas uma caixa delimitadora (*bounding box*), o MTCNN é um modelo de Deep Learning que executa múltiplas tarefas: ele não só classifica uma região da imagem como "rosto" ou "não-rosto", mas também regride a localização de cinco pontos de referência fiduciais: os centros dos dois olhos, as pontas do nariz e os cantos da boca.

O processo de alinhamento utilizou especificamente as coordenadas dos olhos. Para cada imagem, o algoritmo executou os seguintes passos:

1. **Localização dos Pontos Oculares:** As coordenadas (x, y) do olho esquerdo e do olho direito foram extraídas do resultado do MTCNN.
2. **Cálculo do Ângulo Interocular:** Utilizando as coordenadas dos dois pontos, foi calculado o ângulo da linha que os conecta em relação ao eixo horizontal. Este ângulo representa a inclinação ou rotação da cabeça na imagem. A função arctan2(dy, dx) foi utilizada para obter um ângulo preciso, considerando todos os quadrantes.
3. **Correção Rotacional:** Com o ângulo de inclinação determinado, foi aplicada uma **transformação afim** de rotação na imagem inteira. A matriz de rotação foi calculada para girar a imagem em torno do seu centro por um ângulo oposto ao da inclinação. O resultado é uma nova imagem onde a linha que conecta os olhos está perfeitamente na horizontal.

A justificativa para este procedimento é a **normalização da pose**. As redes neurais são sensíveis a variações rotacionais. Ao garantir que todos os rostos no dataset estejam alinhados da mesma forma, removemos a inclinação da cabeça como uma variável de confusão. Isso força o modelo de aprendizado a focar-se exclusivamente nas deformações dos músculos faciais (a microtextura e a geometria da expressão) que são intrínsecas à emoção, em vez de aprender a associar uma emoção a uma determinada pose, o que aumenta significativamente a robustez e a capacidade de generalização do classificador.

2. Extração e Recorte do Rosto: Isolando a Região de Interesse (ROI)

Subsequentemente ao alinhamento, foi realizado o recorte (cropping) da Região de Interesse (ROI), que neste caso é a face. Este procedimento é fundamental para o sucesso de modelos de Deep Learning, pois atua como uma forma de **atenção espacial**, eliminando informações de fundo que são contextualmente irrelevantes para a tarefa de reconhecimento de emoções. Informações como cenário, roupas ou outras pessoas na imagem são consideradas "ruído" e podem interferir

negativamente no processo de aprendizado, levando o modelo a aprender correlações espúrias.

O processo técnico de extração foi executado da seguinte forma:

1. **Re-deteção na Imagem Alinhada:** Após a rotação da imagem, o MTCNN foi aplicado uma segunda vez. Isso é necessário porque a rotação da imagem altera as coordenadas originais da caixa delimitadora. A nova deteção fornece a bounding box precisa na imagem já alinhada.
2. **Obtenção das Coordenadas:** O resultado desta deteção forneceu um conjunto de quatro coordenadas: [x, y, w, h], que definem o canto superior esquerdo do retângulo (x, y), sua largura (w) e sua altura (h).
3. **Operação de Recorte (Slicing):** O recorte em si é uma operação de **fatiamento de matriz (matrix slicing)**. Como as imagens são representadas computacionalmente por matrizes NumPy, a extração do rosto é realizada selecionando um subconjunto desta matriz. A operação imagem_alinhada[y:y+h, x:x+w] seleciona todos os pixels contidos nas linhas entre y e y+h e nas colunas entre x e x+w.

O resultado desta etapa é uma nova imagem, de dimensões menores, que contém exclusivamente os pixels da face alinhada. Ao isolar a ROI, garantimos que 100% do poder computacional e da capacidade de aprendizado do modelo sejam focados na extração de características faciais, como explicado nos conceitos de *Feature Extraction*, maximizando a eficiência e a precisão da classificação.

3. Redimensionamento (Resizing): Todas as imagens de rosto extraídas foram submetidas a uma padronização dimensional para um tamanho fixo de **224x224 pixels**. Esta resolução foi escolhida por ser o padrão de entrada para a maioria das arquiteturas pré-treinadas no dataset ImageNet. A conformidade com este pré-requisito é essencial para a aplicação bem-sucedida do Transfer Learning, permitindo que os modelos aproveitem o conhecimento prévio adquirido.

4. Conversão para Escala de Cinza (Grayscale): As imagens foram convertidas de um espaço de cor RGB de 3 canais para um espaço acromático de 1 canal. Esta foi uma decisão metodológica para simplificar a dimensionalidade dos dados de entrada e forçar o modelo a focar-se nas características estruturais e de textura da face (contornos, rugas, etc.), que são os indicadores primários das emoções básicas. Adicionalmente, esta conversão mitiga a variância introduzida por diferentes condições de iluminação e tons de pele.

5. Balanceamento de Classes (Offline): Mitigação de Viés na Distribuição de Dados

A Análise Exploratória de Dados (EDA) revelou um severo desbalanceamento na distribuição de classes do conjunto de treino original da RAF-DB, uma característica comum em datasets de emoção coletados "in-the-wild". Especificamente, a classe 'Felicidade' continha 4.779 amostras, enquanto a classe 'Medo' possuía apenas 272. Um modelo de Deep Learning treinado diretamente sobre dados com tal desequilíbrio desenvolveria um forte viés, otimizando seu aprendizado para as classes majoritárias em detrimento das minoritárias, resultando em baixa performance e incapacidade de generalização para as emoções menos representadas.

Para mitigar este viés intrínseco aos dados, foi implementada uma **estratégia de balanceamento híbrida e offline**. O objetivo foi gerar um novo conjunto de dados de treino com uma distribuição de classes perfeitamente uniforme, estabelecendo um alvo de **1.000 imagens por cada uma das 7 emoções**. Esta abordagem foi dividida em duas técnicas complementares, aplicadas de acordo com a representatividade de cada classe.

Técnica 1: Subamostragem Aleatória (Undersampling) para Classes Majoritárias

- **Aplicação:** Esta técnica foi aplicada às classes com mais de 1.000 amostras ('Felicidade', 'Neutro', 'Tristeza', 'Surpresa').
- **Procedimento:** Para cada uma dessas classes, foi realizado um processo de amostragem aleatória sem reposição, onde 1.000 instâncias foram selecionadas do conjunto original de imagens já processadas (alinhadas, recortadas e em escala de cinza). As amostras excedentes foram descartadas do conjunto de treino final.
- **Justificativa:** A subamostragem foi escolhida como uma estratégia pragmática para reduzir drasticamente a dominância das classes majoritárias e, como benefício secundário, diminuir o tamanho total do dataset, otimizando a eficiência computacional de cada época de treinamento. Embora esta abordagem incorra no risco de perda de informação — pois amostras potencialmente valiosas são descartadas —, a decisão foi considerada um trade-off aceitável, dado o volume massivo da classe 'Felicidade' e a necessidade de mitigar seu forte viés.

Técnica 2: Sobreamostragem Sintética via Aumento de Dados (Oversampling) para Classes Minoritárias

- **Aplicação:** Esta técnica foi aplicada às classes com menos de 1.000 amostras ('Raiva', 'Nojo', 'Medo').
- **Procedimento:** O objetivo foi aumentar a representação destas classes sem introduzir duplicatas exatas, o que poderia levar ao sobreajuste (*overfitting*). A estratégia consistiu em:
    1. Manter todas as amostras originais da classe no novo conjunto de dados.
    2. Calcular o número de amostras sintéticas necessárias para atingir o alvo de 1.000.
    3. Gerar cada nova amostra selecionando aleatoriamente uma imagem do conjunto original daquela classe e aplicando-lhe um conjunto de **transformações geométricas leves e aleatórias (Data Augmentation)**, como rotações (em um intervalo de -15 a +15 graus) e espelhamento horizontal. Cada imagem aumentada foi salva como um novo ficheiro.
- **Justificativa:** A sobreamostragem garante que nenhuma informação das classes raras seja perdida. O uso de aumento de dados para criar os novos exemplos é crucial, pois introduz uma variância saudável nos dados, forçando o modelo a aprender características mais generalizáveis da expressão, em vez de simplesmente memorizar as poucas amostras originais. Esta abordagem sintética aumenta a "voz" das classes minoritárias de forma robusta.

A aplicação sequencial destas duas técnicas resultou em um conjunto de dados de treino final, perfeitamente balanceado, contendo **7.000 imagens**. Esta abordagem híbrida assegura que o modelo, durante a fase de treinamento, seja exposto a uma distribuição de classes uniforme, forçando-o a dedicar igual importância ao aprendizado das características de todas as sete emoções.

4. Fase 2: Pré-processamento Leve (Transformações Online)

Esta fase de transformações é aplicada dinamicamente em memória durante o ciclo de treinamento do modelo.

Técnicas a serem Aplicadas:

1. **Normalização de Pixels:** Os valores de pixel de cada imagem, no intervalo de inteiros [0, 255], serão escalonados para o intervalo de ponto flutuante [0.0, 1.0]. Esta normalização é uma prática padrão indispensável em Deep Learning, pois garante que os dados de entrada tenham uma escala

pequena e consistente, o que estabiliza e acelera significativamente o processo de convergência durante o treinamento.

5. Resultado Final do Processo

A execução bem-sucedida do pipeline de pré-processamento culminou na geração de um novo dataset otimizado na pasta data/processed/raf_db_balanced/. Este dataset contém:

- Um conjunto de **treino** perfeitamente balanceado, com **7.000 imagens** (1.000 por cada uma das 7 emoções).
- Um conjunto de **teste** com a distribuição original, contendo **3.068 imagens**.

Todas as imagens neste novo dataset estão limpas e padronizadas: são rostos alinhados, em escala de cinza e com resolução de 224x224 pixels, encontrando-se agora em estado ideal para a subsequente fase de treinamento dos modelos.