



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS
CURSO DE GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO

Francisco Vicenzi

**Identificação de propriedades de produtos a partir da leitura automática de
rótulos**

Florianópolis
[2021]

Francisco Vicenzi

Identificação de propriedades de produtos a partir da leitura automática de rótulos

Trabalho de Conclusão de Curso do Curso de Graduação em Ciências da Computação do Campus Florianópolis da Universidade Federal de Santa Catarina para a obtenção do título de bacharel em Ciências da Computação.

Orientador: Prof. Dr. Mauro Roisenberg

Coorientador: Lucas May Petry

Florianópolis

[2021]

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Vicenzi, Francisco

Identificação de propriedades de produtos a partir da
leitura automática de rótulos / Francisco Vicenzi ;
orientador, Mauro Roisenberg, coorientador, Lucas Petry,
2021.

88 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Ciências da Computação, Florianópolis, 2021.

Inclui referências.

1. Ciências da Computação. 2. Reconhecimento Ótico de
Caracteres. 3. Leitura Automática. 4. Inteligência
Artificial. 5. Visão Computacional. I. Roisenberg, Mauro.
II. Petry, Lucas. III. Universidade Federal de Santa
Catarina. Graduação em Ciências da Computação. IV. Título.

Francisco Vicenzi

Identificação de propriedades de produtos a partir da leitura automática de rótulos

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de “bacharel em Ciências da Computação” e aprovado em sua forma final pelo Curso de Graduação em Ciências da Computação.

Florianópolis, 27 de setembro de 2021.

Prof. Dr. Jean Everson Martina
Coordenador do Curso

Banca Examinadora:

Prof. Dr. Mauro Roisenberg
Orientador

Msc. Lucas May Petry.
Avaliador

Prof. Dr. Alexandre Gonçalves Silva.
Avaliador
Universidade Federal de Santa Catarina

A todos que tiveram a paciência de fazer parte.

AGRADECIMENTOS

Uma jornada e tanto. Longa, eu diria, mas necessária. Muitas pessoas passaram; algumas permaneceram. Gostaria de agradecer primeiramente à minha família, que sem o apoio não teria chego até aqui. Mãe, pai e minha irmã, muito obrigado. Além deles, é necessário reconhecer a importância também dos presentes neste projeto; professor Mauro, que mostrou-se um grande orientador, presente e muito compreensivo; Lucas, que aceitou o desafio da coorientação e que fez ótimas contribuições; professor Alexandre, que apontou diversas questões pertinentes como avaliador deste trabalho.

Aos meus amigos, também tenho muito a agradecer. Das primeiras semanas de calouro às vésperas *daquelas* provas. Do EAD até o *spam* no chat da apresentação. Daqui para muito mais, eu espero. Obrigado a todos que participaram.

*“Tanto o autor deste diário quanto o "Diário"
em si são, bem entendido, fictícios.
No entanto, as pessoas que escrevem tais diários
não apenas podem como devem existir
em nossa sociedade, considerando aquelas
circunstâncias em que nossa
sociedade se formou de modo geral.
Eu queria destacar,
tornando-a mais visível para o público,
uma das personalidades do nosso passado recente.
Seria um dos representantes da geração que ainda está viva.”
(Fiódor Dostoiévski, 1864)*

RESUMO

O primeiro contato de um consumidor com o produto é através do rótulo. Os rótulos apresentam as informações essenciais sobre a natureza de um produto e devem, na maior parte dos casos, seguir legislações específicas. Além do consumidor, nota-se neste contexto o papel do produtor, que deve apresentar de maneira correta o rótulo, e a agência reguladora, que tem como função definir e fiscalizar as normas referentes à rotulagem. A identificação de propriedades de produtos a partir da leitura automática de rótulos pode representar um ganho em logística, segurança, acessibilidade e comodidade para todas as partes envolvidas. Deste modo, este projeto apresentará uma aplicação capaz de ler automaticamente rótulos, identificar atributos que sejam úteis e validá-los utilizando-se de regras pré-definidas, tais como restrições alimentares, especificidades de rótulos e valores nutricionais. Para isso, serão utilizadas técnicas de Inteligência Artificial, abrangendo desde a parte de Visão Computacional, através do Reconhecimento Ótico de Caracteres e Aprendizado de Máquina. Além da aplicação proposta, também é apresentada uma revisão bibliográfica do Estado da Arte e métricas para avaliar o desempenho das técnicas utilizadas.

Palavras-chave: reconhecimento ótico de caracteres, leitura automática, aprendizado de máquina, inteligência artificial, visão computacional.

ABSTRACT

The first encounter that a consumer has with a product is through the label. Labels present the essential information about the product and need to follow specific laws. Besides the consumer, one can notice in this context the role of the producer, who must present the label correctly, and the regulatory agency, whose function is to define and supervise the standards for labeling. The identification of product properties from the automatic reading of labels can represent a gain in logistics, safety, accessibility, and convenience for all stakeholders. Thus, this project will present an application that can automatically read labels, identify useful attributes, and validate them using predefined rules, such as dietary restrictions, label specifications, and nutritional values. To accomplish this objective, Artificial Intelligence techniques will be used, ranging from the Computer Vision part, through Optical Character Recognition and Machine Learning. Besides the proposed application, a bibliographic review of the State of the Art and metrics to evaluate the performance of the techniques used are also presented.

Keywords: optical character recognition, automatic reading, machine learning, artificial intelligence, computer vision.

LISTA DE FIGURAS

Figura 1 – Foto de um cachorro dormindo	17
Figura 2 – Representação matricial resumida da imagem à esquerda	17
Figura 3 – Exemplo de aplicação de limiarização	18
Figura 4 – Exemplo de dilatação e erosão	19
Figura 5 – Exemplo de abertura	19
Figura 6 – Exemplo de fechamento	20
Figura 7 – Exemplo de remoção de ruído	20
Figura 8 – Rótulos com ângulo original	25
Figura 9 – Rótulo com ângulo original, aproximado	25
Figura 10 – Rótulo planificado, sobre superfície plana	26
Figura 11 – Componentes extraídos de um rótulo planificado	27
Figura 12 – Componentes extraídos de um rótulo plano	27
Figura 13 – Fluxograma do primeiro método, primeiro e terceiro experimentos . . .	28
Figura 14 – Exemplo de imagem que não produziu resultado no primeiro grupo dos experimentos	29
Figura 15 – Exemplo obtido no primeiro grupo dos experimentos. À esquerda, a imagem original. À direita, a imagem com as caixas de identificação obtidas com o Tesseract	30
Figura 16 – Fluxograma do segundo método, segundo experimento	30
Figura 17 – Exemplo obtido no segundo grupo dos experimentos. À esquerda, a imagem original, em escala de cinza. No meio, a imagem aplicada à abertura e embaçamento. À direita, a imagem aplicada ao Canny. . . .	32
Figura 18 – Fluxograma do terceiro método, quarto experimento	34
Figura 19 – Exemplo de processo do experimento quatro. Na esquerda superior, a imagem original em escala de cinza. Na direita superior, a imagem após MSER e filtragem de regiões. Na parte inferior centralizada, a imagem após junção das regiões.	35
Figura 20 – Fluxograma do quarto método, quinto experimento	37
Figura 21 – Cabeçalho do aplicativo	39
Figura 22 – Definições de processamento de imagem e texto	40
Figura 23 – Exibição das imagens processadas	45
Figura 24 – Apresentação dos resultados	46

LISTA DE TABELAS

Tabela 1 – Exemplo de distâncias computadas entre palavras	22
Tabela 2 – Sumário do conjunto de dados	27
Tabela 3 – Resultado obtido no primeiro grupo de experimentos.	29
Tabela 4 – Resultado de uma imagem do segundo experimento, dividido pelas técnicas aplicadas	31
Tabela 5 – Resultados obtidos no terceiro grupo de experimentos	33
Tabela 6 – Resultados obtidos no quarto grupo de experimentos	36
Tabela 7 – Resultados obtidos no quinto grupo de experimentos	41
Tabela 8 – Resultados obtidos no quinto grupo de experimentos	42
Tabela 9 – Resultados obtido no quinto grupo de experimentos	43
Tabela 10 – Exemplo de recuperação de dígitos no quinto grupo de experimentos	44
Tabela 11 – Sumário quantitativo para comparação dos métodos apresentados. * A etapa de transformação da imagem não foi realizada. Em negrito, destaca-se a melhor acurácia para aquele grupo; sublinhada, a segunda melhor acurácia.	44

SUMÁRIO

1	INTRODUÇÃO	13
1.1	OBJETIVOS	14
1.2	OBJETIVOS ESPECÍFICOS	14
1.3	ESTRUTURA DO TRABALHO	15
2	CONCEITOS BÁSICOS E TRABALHOS RELACIONADOS	16
2.1	CONCEITOS BÁSICOS	16
2.1.1	Aprendizado de Máquina	16
2.1.2	Visão Computacional	16
2.1.2.1	Limiarização	17
2.1.2.2	Operadores morfológicos	18
2.1.2.3	Remoção de ruído	19
2.1.2.4	Análise de regiões	21
2.1.3	Reconhecimento Ótico de Caracteres	21
2.1.3.1	Tesseract	21
2.1.4	Métricas	22
2.1.4.1	Distância de Levenshtein	22
2.1.4.2	Acurácia	22
2.2	TRABALHOS RELACIONADOS	23
3	DESENVOLVIMENTO	25
3.1	CONJUNTO DE DADOS	25
3.2	EXPERIMENTOS	26
3.2.1	Primeiro experimento	28
3.2.1.1	Descrição	28
3.2.1.2	Resultados	28
3.2.2	Segundo experimento	28
3.2.2.1	Descrição	28
3.2.2.2	Resultados	29
3.2.3	Terceiro experimento	30
3.2.3.1	Descrição	30
3.2.3.2	Resultados	32
3.2.4	Quarto experimento	32
3.2.4.1	Descrição	32
3.2.4.2	Resultados	34
3.2.5	Quinto experimento	34
3.2.5.1	Descrição	34
3.2.5.2	Resultados	36
3.3	SUMÁRIO QUANTITATIVO DOS RESULTADOS	38

3.4	APLICAÇÃO	38
4	CONCLUSÃO	47
	REFERÊNCIAS	48
	APÊNDICE A – CÓDIGO FONTE	50
	APÊNDICE B – ARTIGO	63

1 INTRODUÇÃO

Rótulos correspondem, em sua maioria, à fonte primária de informação acerca de um produto que um consumidor pode ter. Em relação a produtos alimentícios, ingredientes utilizados, valores nutricionais, prazo de validade são exemplos de informações apresentadas, geralmente, em rótulos. Como a legislação sobre rotulagem varia entre países, é necessário estar atento ao padrão nacional a ser utilizado. No Brasil, o órgão que estabelece e fiscaliza as normas de rotulagem é a Anvisa (AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA - ANVISA, 2020). Deste modo, são definidas, além das informações obrigatórias, maneiras de como dispô-las na impressão do rótulo.

Podemos identificar três agentes principais interessados na rotulagem correta: consumidor, produtor e agência reguladora. É de interesse do consumidor que as informações estejam claramente disponíveis, arranjadas de uma maneira que seja fácil identificar questões importantes para a sua saúde, desde ingredientes que possam ser alergênicos até valores nutricionais específicos para uma dieta. O papel da agência reguladora consiste, em sua essência, estabelecer normas e fiscalizar produtos para que, no mínimo, o consumidor tenha ciência do que esteja consumindo. Já o produtor presta atenção na rotulagem justamente para seguir as normas previstas, evitando processos judiciais e fomentando uma clientela fidelizada, a partir da qualidade e coerência de seus produtos.

A identificação de propriedades de produtos a partir da leitura automática de rótulos pode representar um ganho para as três partes citadas anteriormente. Aliado a um conjunto de regras, podemos estabelecer e abstrair diversas tarefas a serem realizadas por essa aplicação. Por exemplo, poderia-se carregar as normas da Anvisa e checar, produto por produto, se os rótulos possuem todas as informações necessárias para a comercialização, se estão danificados, etc. Ainda, um consumidor poderia, com seu *smartphone*, tirar uma foto de um produto e verificar se os valores nutricionais correspondem para sua dieta, se possui algum ingrediente que deva prestar atenção, etc. Em relação à acessibilidade, seria possível aliar à aplicação um processo de transcrição de texto para fala, permitindo que pessoas com baixa visão possam ter um maior entendimento do produto em suas mãos.

Neste projeto, utilizaremos técnicas de Visão Computacional (CV) e Aprendizado de Máquina (ML) para desenvolver uma aplicação capaz de identificar propriedades de imagens de rótulos. A identificação automática de propriedade de textos corresponde a um problema complexo (GONZALEZ *et al.*, 2012) e recorrente do campo da Visão Computacional. Fatores como diferentes tipos de fontes, cores, tamanhos, ângulos e distorções de imagem, por exemplo, tornam o reconhecimento de texto através de imagens uma tarefa difícil. Neste interím, para a primeira parte da implementação, utilizaremos técnicas de Reconhecimento Ótico de Caracteres (OCR).

O OCR corresponde ao processo de identificar padrões em imagens digitalizadas, fazendo assim a sua correspondência para caracteres pré-definidos (alfanuméricos ou não)

(CHAUDHURI *et al.*, 2017). Existem diversas maneiras de se realizar OCR. Entretanto, alguns passos costumam ser comuns aos diversos tipos de projeto: varredura, ou *scan*, da imagem, à procura dos textos; segmentação dos caracteres; extração de *features*; treino dos modelos e classificação dos caracteres; pós-processamento (SINGH, 2013) (CHAUDHURI *et al.*, 2017). Para etapa de treinamento e classificação, será utilizada uma *engine* de código aberto de OCR, chamada Tesseract. A etapa de pós-processamento consistirá na validação das informações obtidas para com as regras previamente estabelecidas para o sistema, sejam elas legislações, restrições alimentares, etc. Nesta etapa, ainda é possível a utilização de Modelos de Linguagem, técnica esta de Processamento de Linguagem Natural, para melhorar a performance da nossa aplicação. Modelos de Linguagem, como o BERT (DEVLIN *et al.*, 2018), podem ser ferramentas importantes para minimizar erros oriundos do OCR (NGUYEN *et al.*, 2020) ou, até mesmo, apontar erros do próprio rótulo. Estratégias menos robustas, tal como similaridade de texto, também podem ser utilizadas para o pós-processamento.

Portanto, a leitura de rótulos de produtos, ainda que trivial para humanos, consiste em um desafio tecnológico a ser superado. Sua automatização pode trazer diversos benefícios, seja de logística, segurança, acessibilidade ou comodidade. Ao fim deste projeto, apresentaremos uma aplicação que realize a leitura automática de rótulos, identificando atributos que sejam úteis para classificação e validação através de regras pré-definidas.

1.1 OBJETIVOS

O objetivo principal do projeto é desenvolver uma aplicação que, dada uma imagem de rótulo, seja capaz de realizar a leitura dos textos, identificando, automaticamente, propriedades do produto em questão.

1.2 OBJETIVOS ESPECÍFICOS

1. cruzar propriedades identificadas a partir da leitura dos rótulos com uma base de regras pré-definidas, a fim de responder questões sobre o conteúdo do rótulo;
2. pesquisar, utilizar e aprimorar métodos de Reconhecimento Ótico de Caracteres;
3. avaliar e utilizar técnicas de Visão Computacional para o pré-processamento de imagens para melhorar resultados de identificação de propriedades;
4. criar um protótipo de aplicação que utilize dos conhecimentos obtidos neste projeto;
5. disponibilizar código-fonte da aplicação de maneira acessível ao público interessado.

1.3 ESTRUTURA DO TRABALHO

Além do presente capítulo, que trata da introdução e apresentação do projeto, este trabalho está estruturado em três capítulos, a serem explicados a seguir.

O capítulo 2 apresenta os conceitos básicos para o entendimento do trabalho e os trabalhos relacionados. De uma forma superficial, são introduzidos os conceitos de aprendizado de máquina, visão computacional e reconhecimento ótico de caracteres. Ainda em relação à visão computacional, conceitos de limiarização, operadores morfológicos e análises de regiões são apresentados como métodos de processamento de imagens. Além disso, também são apresentados trabalhos que tenham relação com o tema do projeto.

No capítulo 3, o desenvolvimento do trabalho é apresentado. Ele foi dividido em duas partes. Na primeira parte do capítulo, é feita uma exposição do conjunto de dados produzido e utilizado. Já na segunda parte, são apresentados os experimentos realizados. Cada experimento foi dividido em duas partes: descrição e resultados.

O capítulo 4 corresponde à conclusão do trabalho. Nele, é discutido o que foi feito e apresentados os trabalhos futuros, a serem realizados na próxima etapa do projeto.

2 CONCEITOS BÁSICOS E TRABALHOS RELACIONADOS

2.1 CONCEITOS BÁSICOS

2.1.1 Aprendizado de Máquina

A invenção das máquinas, de um modo geral, foi impulsionada pelo desejo de facilitar a vida humana, tornando mais simples diversos processos. Com a evolução tecnológica, muitas tarefas, que até então eram complexas e demandavam muito tempo, tornaram-se simples a partir da utilização da computação. Por exemplo, podemos citar o cálculo, com as suas mais diversas aplicações.

Para máquinas, tarefas como operações matemáticas são relativamente simples, enquanto humanos apresentam uma maior dificuldade. Entretanto, existem diversos problemas que não são facilmente descritos e resolvidos computacionalmente. Desse modo, o Aprendizado de Máquina surge como um meio de lidar com essa problemática. Ao invés de serem explicitamente programados para uma tarefa, esses sistemas aprendem através dos dados de entrada fornecidos (GOODFELLOW; BENGIO; COURVILLE, 2016).

Diversas tarefas podem ser trabalhadas através do aprendizado de máquina. Problemas de regressão, como para a estimação de tempo de estadia em hospitais, e problemas de classificação, como para prever a qual classe um animal pertence, são exemplos dessas tarefas.

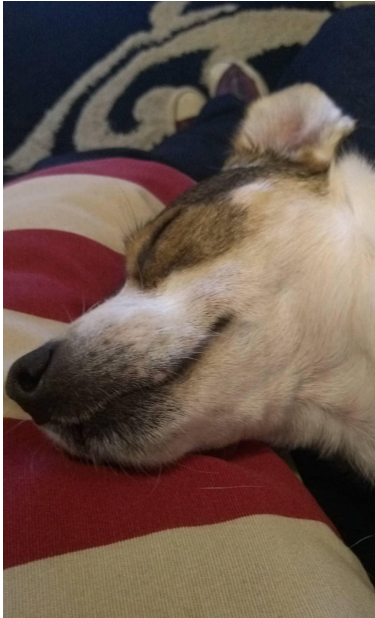
2.1.2 Visão Computacional

A visão computacional representa a área de estudo na qual o principal interesse consiste em desenvolver sistemas e técnicas capazes de obter informações úteis de imagens (JAIN *et al.*, 1995). O ato de observar uma cena e entender o que se passa nela é uma tarefa simples, em geral, para humanos. Desde crianças, é natural que desenvolvamos a capacidade de reconhecer padrões e identificar elementos que façam parte da composição do ambiente. Entretanto, traduzir essa capacidade para computadores é uma tarefa extremamente difícil e que representa décadas de estudo até então.

A figura 1 apresenta uma foto tirada pelo autor em 2019. Pela ótica de um ser humano, facilmente conseguimos descrever a cena: mais à frente, um cachorro dormindo em uma almofada com listras brancas e vermelhas; ao fundo, desfocado, um par de tênis sobre um tapete azul e branco. Já pela ótica de um computador, porém, o que é possível recuperar de informação?

Enquanto a visão humana acontece em três dimensões, as imagens são representadas digitalmente em apenas duas dimensões. Em suma, o objeto de trabalho da visão computacional corresponde a matrizes de pixels que representam imagens. Por exemplo, a figura 2 apresenta a forma matricial resumida da imagem 1, onde cada elemento corresponde aos valores RGB de cada pixel. Também é possível representar imagens de formas

mais simples, em que as mais comuns são em escala de cinza e de forma binária. Na escala de cinza, utiliza-se apenas um valor por pixel, que varia de 0 a 255; em imagens binárias, os pixels podem assumir apenas dois valores, 0 e 1.



$$\begin{bmatrix} [32, 42, 67] & [30, 40, 65] & \cdots & [48, 46, 34] \\ [31, 41, 66] & [29, 39, 64] & \cdots & [43, 41, 29] \\ \vdots & \vdots & \vdots & \vdots \\ [131, 119, 97] & [133, 121, 99] & \cdots & [123, 111, 89] \\ [132, 120, 98] & [133, 121, 99] & \cdots & [122, 110, 88] \end{bmatrix}$$

Figura 2 – Representação matricial resumida da imagem à esquerda

Figura 1 – Foto de um cachorro dormindo

Portanto, os métodos de visão computacional, ao agir sobre as representações digitais de imagens, almejam, de certa forma, reconstruir as propriedades do mundo real, a fim de descrevê-lo (SZELISKI, 2010). Apesar de todas as dificuldades, diversas técnicas já são utilizadas no cotidiano, como, por exemplo, na leitura automática de placas de automóveis; na telemedicina, reconhecimento de objetos, direção autônoma de automóveis, etc.

2.1.2.1 Limiarização

Em muitos casos, imagens podem possuir certas propriedades que dificultem técnicas de processamento. Desse modo, separar partes relevantes da imagem pode ser benéfico (ou até mesmo o resultado esperado). Uma das técnicas que auxilia nessa questão é a limiarização. A limiarização consiste em selecionar os valores de intensidade de uma imagem a partir de um valor de limiar definido (JAIN *et al.*, 1995).

A figura 3 apresenta um exemplo de imagem em que a limiarização foi aplicada, a fim de representá-la de forma binária. Neste exemplo, foi utilizado um limiar de 75, selecionado arbitrariamente. Entretanto, não necessariamente este é o melhor valor para realizar a limiarização. Existem técnicas de obter automaticamente os valores de limiar, onde as características da imagem são levadas em consideração na decisão.

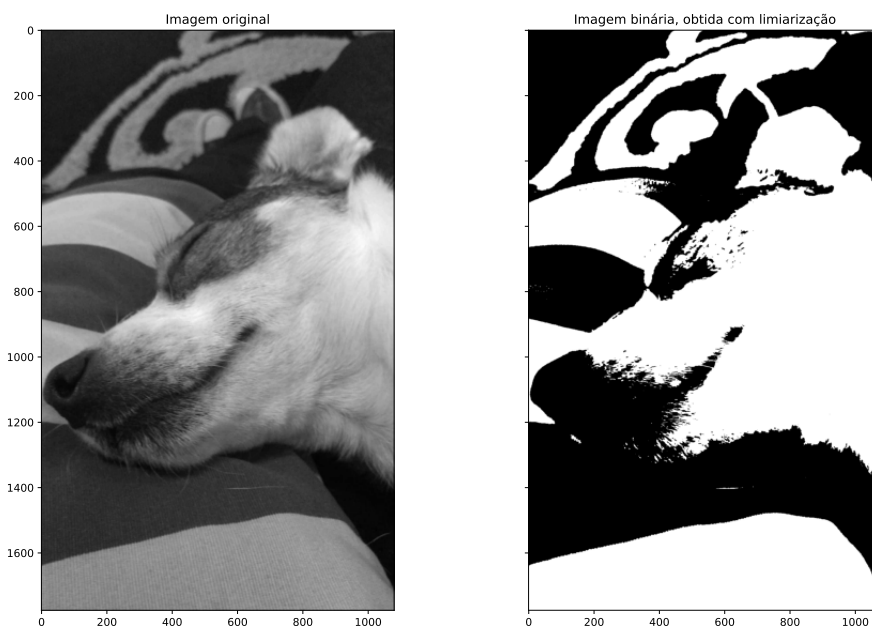


Figura 3 – Exemplo de aplicação de limiarização

2.1.2.2 Operadores morfológicos

Em muitos casos, pensar a partir da abordagem da forma de objetos pode facilitar o entendimento e desenvolvimento de técnicas de processamento de imagens (JAIN *et al.*, 1995). Desse modo, os operadores morfológicos são utilizados para obter, modificar e representar formas em imagens binárias.

No processamento morfológico, estruturas matemáticas chamadas de elementos estruturantes ou *kernel* são utilizadas para realizar as operações nas imagens. Essas estruturas podem assumir formas variadas, desde simples matrizes 3×3 compostas apenas por números 1 até formatos como cruzes, elipses, entre outras (SZELISKI, 2010). Desse modo, o kernel escolhido é utilizado para realizar o escaneamento dos pixels da imagem binária. A operação morfológica consiste na escolha de quais e/ou quantos valores serão considerados do elemento estruturante, a cada etapa.

As operações morfológicas mais comuns são a dilatação, erosão, abertura e fechamento. Na dilatação, sempre que o valor central do kernel coincidir com o pixel da imagem, os valores não-nulos do elemento estruturante são copiados. Já erosão representa a operação complementar; os valores da imagem são mantidos apenas quando os valores coincidirem totalmente com o kernel utilizado. A abertura corresponde à uma erosão seguida por dilatação, enquanto o fechamento é feito a partir de uma dilatação seguida por erosão.

A figura 4 apresenta um exemplo das operações de dilatação e erosão, utilizando um kernel 7×7 . Em comparação à imagem original, observamos que a imagem foi, de fato, dilatada, enquanto na erosão, a imagem diminuiu. Na figura 5, observamos que, após a abertura, ruídos da imagem foram removidos. Já na figura 6, a operação de fechamento

faz com que buracos presentes na imagem sejam preenchidos.



Figura 4 – Exemplo de dilatação e erosão

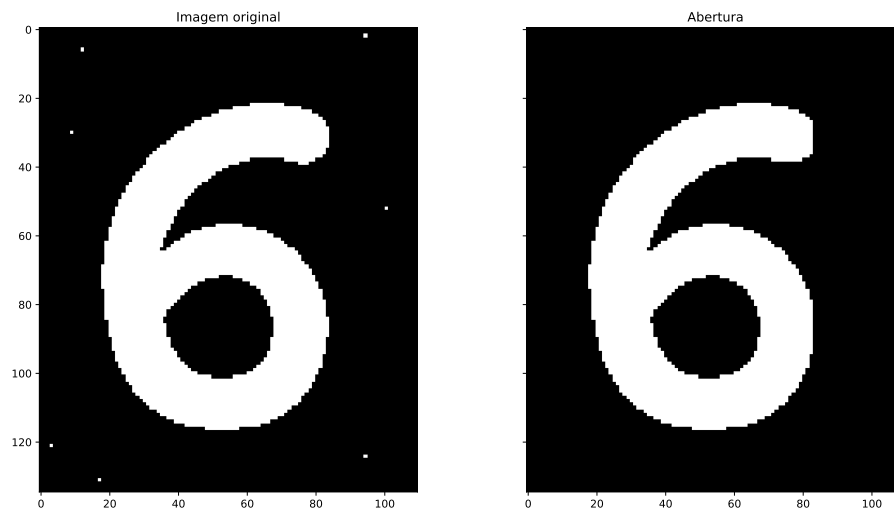


Figura 5 – Exemplo de abertura

2.1.2.3 Remoção de ruído

Por mais que haja muito cuidado no processo de obtenção de imagens, não é raro que diversos ruídos apareçam. Os ruídos, no contexto de imagens digitais, podem variar muito, especialmente de acordo com o que será feito com a imagem em questão. Desse modo, fatores como variação de iluminação, contraste, intensidade e outras características indesejadas devem ser trabalhadas na etapa de pré-processamento (JAIN *et al.*, 1995). Duas técnicas que exemplificam esse cuidado são a equalização de histograma e a suavização Gaussiana.

A suavização Gaussiana consiste em uma técnica de suavização que utiliza a função Gaussiana na escolha dos pesos a serem aplicados na filtragem linear de uma imagem (JAIN *et al.*, 1995). Os filtros Gaussianos são amplamente utilizados pois atuam como

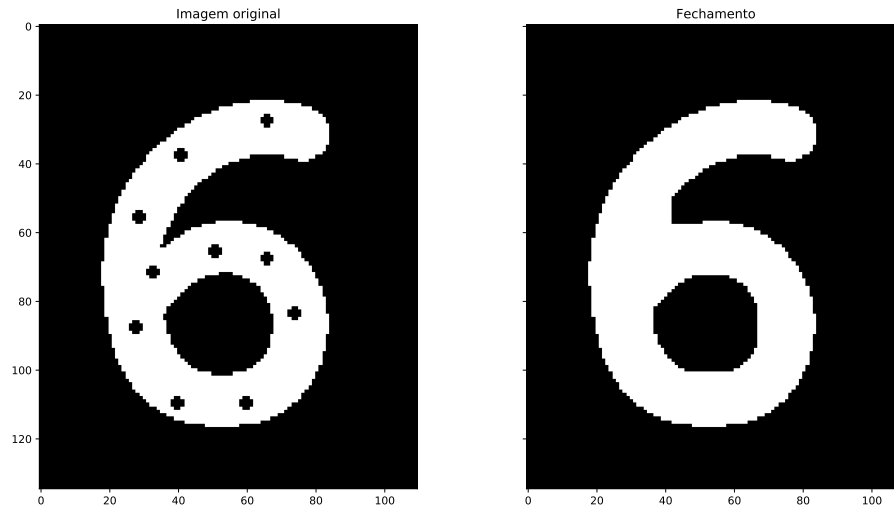


Figura 6 – Exemplo de fechamento

filtros de passa-baixo tanto nos domínios de espaço e frequência. Uma das principais aplicações desse tipo de suavização consiste na remoção de ruídos do tipo sal-e-pimenta e gaussiano. Esses ruídos consistem em ocorrências aleatórias de valores de intensidade preto e branco, que dificultam tanto a visualização como o processamento da imagem.

A equalização de histograma é uma técnica que tem como objetivo ajustar o contraste de imagem com base no seu histograma de intensidade. Desse modo, busca-se redistribuir uniformemente os valores de cinza (JAIN *et al.*, 1995). O ajuste do contraste da imagem pode ser benéfico para a visualização humana e para o realce de regiões, cenários importantes para as etapas de pré-processamento.



Figura 7 – Exemplo de remoção de ruído

A figura 7 apresenta um exemplo de aplicação de suavização Gaussiana e de equalização de histograma. À esquerda, a imagem original, em escala de cinza, onde observa-se diversos tipos de ruídos, tanto de iluminação como de intensidade. A imagem do meio representa a aplicação da suavização Gaussiana; nela, verificamos que, a partir do borra-

mento, algumas variações de intensidade foram removidas. Isso pode ser útil para alguma determinada *pipeline* de processamento. À direita, por fim, a imagem aplicada à equalização de histograma. Observa-se algumas características interessantes que foram realçadas neste procedimento; tanto o texto da capa como os contornos do cenário estão mais nítidos, com suas formas mais bem definidas.

2.1.2.4 Análise de regiões

Regiões, no contexto de imagens, são conjuntos de pixels que estejam conectados e possuem características similares. A análise de regiões representa uma área importante na visão computacional, uma vez que a sua divisão em partes menores pode possibilitar uma melhor interpretação dos elementos representados (JAIN *et al.*, 1995). A identificação de regiões passa, também, pela identificação de bordas.

A segmentação de regiões é uma área bastante desenvolvida na literatura de visão computacional. Diversos métodos foram desenvolvidos até então, cada qual com suas características (SZELISKI, 2010). Um exemplo consiste no método MSER (MATAS *et al.*, 2004) (acrônimo de *Maximally Stable Extremal Regions*), o qual apresenta uma maneira de extrair regiões de uma imagem baseado na correspondência de seus elementos. Além desse, outros exemplos de métodos de segmentação de regiões variam desde a aplicação de divisão e junção de regiões até a utilização de histogramas de intensidade.

2.1.3 Reconhecimento Ótico de Caracteres

O Reconhecimento Ótico de Caracteres é uma tarefa de visão computacional que consiste na identificação e recuperação de texto em imagens digitais. Portanto, sistemas de OCR (acrônimo de *Optical Character Recognition*) realizam a conversão de caracteres dispostos em imagens para textos digitais. Essas imagens, por exemplo, podem conter caracteres escritos à mão, inscritos ou impressos (SINGH; BACCHUWAR; BHASIN, 2012).

2.1.3.1 Tesseract

O Tesseract é um sistema de Reconhecimento Ótico de Caracteres de código aberto ¹, distribuído a partir da licença Apache 2.0 e financiado pela Google. Atualmente encontra-se na versão 4.1.1 (TESSERACT, 2021). Está disponível para várias plataformas e funciona a partir de linha de comando. Além disso, contém APIs definidas para utilização de aplicativos com sistema gráfico ou bibliotecas.

Até a versão 4.0, a arquitetura consistia, principalmente, na análise de componentes conexos, localização de linhas e reconhecimento de palavras (PATEL; PATEL; PATEL, 2012). A partir da versão 4.0, foi adicionado à arquitetura um sistema baseado em redes neurais LSTM, além de modelos treinados para mais de 120 línguas. Essa mudança

¹ Repositório principal do projeto: <https://github.com/tesseract-ocr/tesseract>

apresentou ganhos fortes em acurácia para o sistema de OCR. Desse modo, é possível utilizar os modelos treinados, aprimorá-los ou, até mesmo, treinar modelos próprios.

2.1.4 Métricas

Duas métricas principais são utilizadas para a avaliação quantitativa dos experimentos: distância de Levenshtein (seção 2.1.4.1) e acurácia (seção 2.1.4.2).

O reconhecimento ótico de caracteres muitas vezes retorna palavras com erros de ortografia. Em função disso, faz-se necessário definir uma maneira de conferir se uma palavra foi recuperada ou não, mesmo que não seja uma combinação perfeita. Desse modo, a distância de Levenshtein é utilizada para a obtenção de um valor numérico de similaridade entre termos. Com isso, é possível definir um limiar de aceitação para possíveis erros de ortografia oriundos do OCR.

A acurácia é uma métrica que fornece um significado numérico para a realização do experimento. Desse modo, podemos utilizá-la, neste contexto, para verificar o desempenho das técnicas aplicadas.

2.1.4.1 Distância de Levenshtein

A distância de Levenshtein consiste em uma métrica de similaridade entre sequências de caracteres. A distância é computada como sendo o menor número de substituições necessárias para transformar uma sequência A em B (KRUSKAL, 1983). O algoritmo é definido a partir de três operações fundamentais: substituição, inserção e remoção.

A tabela 1 mostra exemplos de distâncias computadas entre os termos "batata", "batuta" e "banana". A diagonal principal mostra que a distância entre palavras iguais é nula. A distância entre "batata" e "batuta" é igual a 1; uma substituição é necessária para transformar a primeira na segunda. A maior distância encontrada no exemplo dá-se entre os termos "banana" e "batuta". Para esse caso, são necessárias três substituições.

	batata	batuta	banana
batata	0	1	2
batuta	1	0	3
banana	2	3	0

Tabela 1 – Exemplo de distâncias computadas entre palavras

2.1.4.2 Acurácia

Definição 2.1.1 Acurácia. Dado uma lista de termos recuperados R e uma lista de palavras-chave K , a acurácia é computada como

$$Acc(R, K) = \frac{\#R}{\#K}.$$

2.2 TRABALHOS RELACIONADOS

A identificação automática de texto em imagens é um tema bastante estudado na literatura. Como a leitura ótica de caracteres já é discutida há décadas, uma das maneiras de se buscar melhores resultados é trabalhar no pré-processamento da imagem para, posteriormente, aplicar métodos já existentes de OCR.

Em geral, os algoritmos de reconhecimento de texto, quando utilizados de forma generalizada, podem ter problemas ao lidar com elementos distintos nas imagens, que não sejam caracteres. Assim, Gonzalez *et al.* (2012) apresenta um método de auxílio ao reconhecimento ótico de caracteres. O método proposto abrange a etapa anterior ao reconhecimento: a localização de texto. A segmentação do texto do restante da imagem faz com que a qualidade do processamento e do reconhecimento do texto seja superior à apenas executar os métodos de OCR indiscriminadamente. A metodologia proposta é composta por três etapas: segmentação da imagem para possíveis caracteres; análise baseada em componentes conectados para geração de *features*; e, por fim, classificação de linhas de texto. Na primeira etapa, o algoritmo MSER é utilizado para a identificação de regiões prováveis de se conter textos. São geradas 8 features na segunda etapa do método, como, por exemplo, a proporção e a taxa de ocupação da imagem. A terceira etapa consiste no agrupamento dos caracteres em linhas, utilizando-se de algoritmos de limiarização adaptativos.

Enquanto técnicas mais gerais são desenvolvidas, a popularidade e qualidade do motor de OCR Tesseract fez com que métodos fossem desenvolvidos direcionados a ele. O método proposto por Sporici, Cuşnir e Boianjiu (2020) tem como objetivo melhorar o desempenho do Tesseract a partir do pré-processamento da imagem. Uma das principais contribuições do método é mostrar uma forma de contornar as limitações conhecidas do motor, tais como lidar com possíveis imperfeições da imagem, que fazem com que a qualidade do resultado decaia muito. Neste método, o pré-processamento da imagem é realizado a partir da geração de kernels convolução, de maneira não supervisionada. Desse modo, são realizados ciclos de treino, onde a imagem com o kernel aplicado é levada ao Tesseract e obtido uma pontuação. Os resultados mostraram-se bastante promissores, sendo que os autores apresentaram métricas como *Character Error Rate* (CER), *Word Error Rate* (WER), *Character Accuracy* (CAR), *Precision*, *Recall* e *F1 Score*.

O método proposto por Khasgiwala (2018) realiza a identificação de rótulos nutricionais em embalagens de alimentos utilizando Redes Neurais Convolucionais. Neste trabalho, cada linha do rótulo é segmentada utilizando o algoritmo *Text Skew Angle Wavelets* (TSAW). Desse modo, busca-se extrair cada elemento do rótulo e seu valor associado. Posteriormente, cada linha é levada à Rede Neural Convolucional para identificar o elemento segmentado. Além disso, o Tesseract também é utilizado para realizar o OCR de cada segmento, a fim de comparar os resultados. Um dos pontos importantes desse trabalho consiste na segmentação das linhas do rótulo, assim como a possibilidade

de receber imagens que sejam rotacionadas. Os resultados reportam acurácia máxima de aproximadamente 69%. Nesse sentido, são descritas como razões para a baixa acurácia o tamanho do conjunto de dados, ruído no conjunto de dados, dimensão das imagens, entre outras.

Matsunaga e Sullivan (2015) apresentam uma arquitetura de trabalho baseada em três módulos para a extração de informação nutricional de rótulos de alimentos: pré-processamento da imagem, OCR e pós-processamento do texto. No pré-processamento da imagem, são utilizadas técnicas visão computacional clássica a fim de extrair o rótulo do restante da imagem. Para isso, utiliza-se suavização, aumento do contraste, filtragem e limiarização para preparar a extração. Essas etapas permitem uma melhor identificação do contorno do rótulo, útil para a extração. O rótulo extraído é aplicado ao Tesseract; sua saída é, posteriormente, tratada com algumas técnicas de correção de ortografia. No final, são obtidos pares de termos e seu valor nutricional. A acurácia deste método consiste na porcentagem de termos nutricionais corretamente identificados no conjunto de dados. O método foi testado em conjuntos de dados com características diferentes, como, por exemplo, rotação, cor e iluminação. A maior parte dos testes teve uma acurácia de aproximadamente 80%.

Gundimeda *et al.* (2019) propõe um sistema de extração automática de elementos de rótulo de alimentos baseado em três subsistemas: remoção de fundo da imagem, classificação automática da qualidade da imagem e extração de atributos. O subsistema de remoção de fundo da imagem é composto por técnicas de processamento de imagem para a correção de iluminação, limiarização e projeção. A classificação automática da qualidade da imagem é composta pela classificação entre duas métricas desse subsistema. A primeira é obtida através da detecção de texto de uma área de interesse utilizando análise de regiões e classificador baseado em regras; a segunda é obtida através do OCR da área de interesse. Posteriormente, essas métricas são analisadas para identificar se a imagem é aceitável, rejeitável ou necessita intervenção manual. O subsistema de extração de atributos busca recuperar elementos como o nome da marca, o título do produto, informações nutricionais, entre outras. Todas as etapas de extração passam pela segmentação da imagem e a aplicação de OCR. A identificação de palavras-chave determina qual fluxo de execução para extração de atributo será utilizado, além de definir a necessidade de pós-processamento do texto obtido. Os resultados consolidados do sistema proposto chegaram, para cada extração de atributo, em uma acurácia de mais de 95%.

3 DESENVOLVIMENTO

3.1 CONJUNTO DE DADOS

Para a realização dos experimentos, foi confeccionado um conjunto de dados de imagens de garrafas de água mineral e de produtos alimentícios. Em busca de uma maior diversidade, foram selecionadas quatro marcas diferentes de água mineral e outros três produtos alimentícios. Ainda assim, as imagens foram separadas em quatro grupos para os experimentos, dadas características como resolução e ângulo das fotos. Esses grupos foram nomeados em ordem alfabética, de A a D, com intuito de diferenciar sua nomenclatura dos experimentos.

O grupo A do conjunto de dados é composto por imagens que capturam o rótulo junto ao formato original das garrafas. Fazem parte deste primeiro grupo imagens com o corpo inteiro da garrafa, assim como também alguns recortes mais focados no rótulo. A figura 8 mostra três exemplos de imagens para o primeiro caso, enquanto a figura 9 apresenta um exemplo para o segundo.



Figura 8 – Rótulos com ângulo original

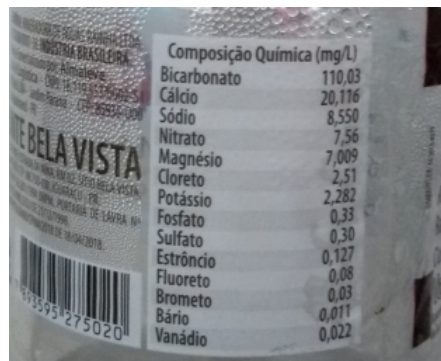


Figura 9 – Rótulo com ângulo original, aproximado

Para o grupo B do conjunto de dados, os rótulos foram extraídos das garrafas e colocados sob uma superfície plana. Desse modo, a relação de ângulo é modificada, uma

vez que perde-se a especificidade do formato do objeto. Na figura 10 é possível observar um exemplo deste grupo.



Figura 10 – Rótulo planificado, sobre superfície plana

Para o grupo C do conjunto de dados, foram extraídos componentes menores a partir das figuras do segundo conjunto de dados. A figura 11 apresenta dois componentes recortados da figura 10.

O grupo D de conjunto de dados é composto por rótulos de três produtos alimentícios diferentes e de algumas imagens de água do terceiro conjunto de dados. Os produtos selecionados foram leite, creme de leite e erva-mate. A partir das fotos, os rótulos foram recortados manualmente de forma digital. A figura 12 apresenta duas imagens presentes nesse conjunto.

O sumário do conjunto de dados é apresentado na tabela 2, dividido pelos grupos, junto às suas características previamente discutidas e suas respectivas quantidades.

3.2 EXPERIMENTOS

Os experimentos realizados nesta etapa tiveram caráter exploratório e de análise quantitativa. Foram realizados em um computador com 4 núcleos i7-75000 CPU 2.70GHZ,



Figura 11 – Componentes extraídos de um rótulo planificado



Figura 12 – Componentes extraídos de um rótulo plano

Grupo	Características	Quantidade
A	Corpo inteiro da garrafa, ângulo e formato original mantidos, grande resolução	44
B	Rótulos planificados, grande resolução	26
C	Recortes de elementos dos rótulos planificados, baixa resolução	13
D	Recortes de elementos dos rótulos planificados, grande resolução	10

Tabela 2 – Sumário do conjunto de dados

com 7.6 Gib de memória RAM, placa de vídeo NVIDIA GeForce GT 620M e sistema operacional Manjaro, na versão do kernel 5.9. O código dos experimentos foi implementado em linguagem Python (v. 3.9), com auxílio principalmente das bibliotecas opencv-python (v. 4.5.1), pytesseract (v. 0.3.7) e numpy (v. 1.19.4). Para capturar a ordem cronológica e curva de aprendizado, dividiu-se os experimentos em cinco grupos principais.

3.2.1 Primeiro experimento

3.2.1.1 Descrição

No primeiro experimento, o objetivo foi a familiarização com o Tesseract, além de buscar entender melhor o problema a ser resolvido. Pode-se dizer que foram realizados testes iniciais exploratórios, uma vez que as imagens foram aplicadas naturalmente, sem nenhum tratamento, no sistema de reconhecimento ótico. Além disso, as imagens utilizadas neste momento foram as do grupo A do conjunto de dados. A ideia principal consistiu em observar qual seria a resposta neste caso, conferir como seria a performance do sistema com as características apresentadas pelas imagens. A figura 13 apresenta um fluxograma que sumariza o primeiro método experimentado.

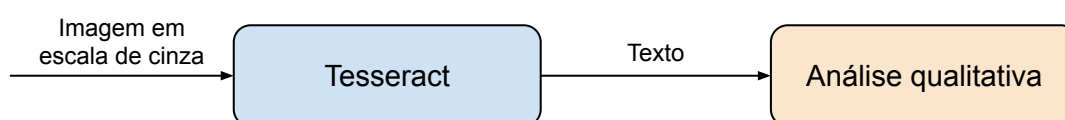


Figura 13 – Fluxograma do primeiro método, primeiro e terceiro experimentos

3.2.1.2 Resultados

Os primeiros resultados não foram satisfatórios, mas serviram como base para pensarmos os próximos experimentos. Em diversas imagens, não foi possível recuperar nenhum texto, apenas algumas letras (ou combinações) sem muito sentido. Uma das imagens que representa este caso é mostrada na figura 14. Entretanto, algumas imagens foram capazes de produzir alguns resultados iniciais, ainda que um número reduzido delas. Uma das maneiras de se analisar a performance, neste momento, foi utilizar uma função do Tesseract para traçar as caixas onde ele identificou letras e palavras. A figura 15 apresenta um exemplo em que foi possível observar um resultado interessante. Na esquerda, temos a imagem original, enquanto na direita é marcada pelas caixas de identificação. Em termos de texto, o resultado obtido é apresentado na tabela 3.

3.2.2 Segundo experimento

3.2.2.1 Descrição

O segundo grupo de experimentos também foi realizado com o primeiro conjunto (A) de dados. Diferente do primeiro experimento, neste momento tentou-se realizar algumas etapas de pré-processamento nas imagens antes de aplicá-las ao Tesseract. As técnicas utilizadas foram de abertura morfológica, remoção de ruído utilizando filtro de mediana, limiarização e Canny.

Imagem	Resultado
Original, em escala de cinza	WOCONTEM (COMPOSIGAO (CARACTERGTERY GLUTEN. QUIMICA (mg/L) FISICO-UIMICAS Estrineo..... 0,037 CAICIO s, 25 : Magnésio..... 0,09 g POtS8I...vc... 0,21 [S0010....core. 75,81 "Vanédo.....,.....mW Sulfato.....fi,28' e 81,10

Tabela 3 – Resultado obtido no primeiro grupo de experimentos.



Figura 14 – Exemplo de imagem que não produziu resultado no primeiro grupo dos experimentos

Uma série de experimentos foi realizada, combinando as técnicas e experimentando alguns parâmetros. As etapas mencionadas são apresentadas na figura 16.

3.2.2.2 Resultados

A experimentação em relação à variação dos parâmetros utilizados pelas técnicas mencionadas não foi muito extensa. Após alguns testes, a configuração deste segundo experimento foi a seguinte: para aplicação da técnica de abertura morfológica, utilizou-se um kernel de 5×5 , composta apenas por números 1; para a remoção de ruído, foi utilizado um filtro mediano de 5×5 ; já para a limiarização, utilizou-se o método de Otsu, que adapta o limiar conforme a imagem; para Canny, os valores de limiar foram 10 e 200.

No geral, como os parâmetros não foram escolhidos de forma tão criteriosa, não

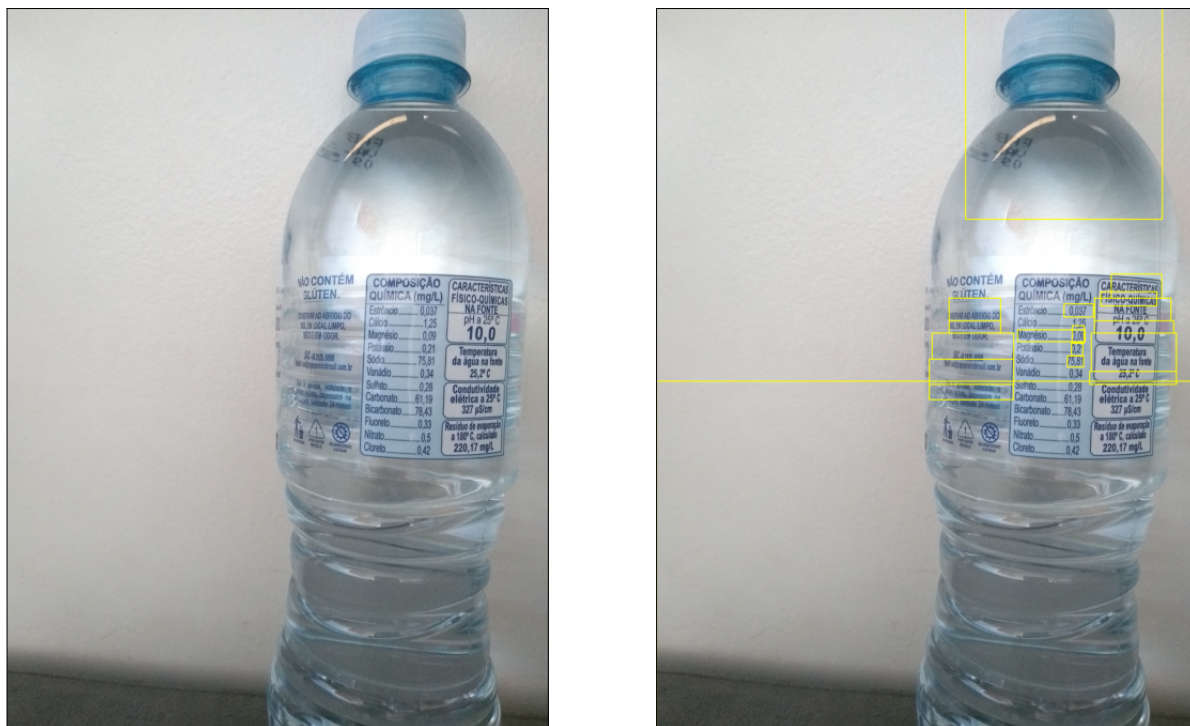


Figura 15 – Exemplo obtido no primeiro grupo dos experimentos. À esquerda, a imagem original. À direita, a imagem com as caixas de identificação obtidas com o Tesseract



Figura 16 – Fluxograma do segundo método, segundo experimento

obteve-se resultados concretos, em geral, neste experimento. Entretanto, a figura 17 apresenta um experimento que produziu resultados interessantes. Vale notar que a imagem original é a mesma da figura 15. Os resultados para esse exemplo estão apresentados na tabela 4. Nota-se que não houve melhora no resultado com as técnicas obtidas; de fato, foram até piores do que com a imagem original, apenas em escala de cinza.

3.2.3 Terceiro experimento

3.2.3.1 Descrição

O terceiro grupo de experimentos foi realizado com o terceiro grupo (C) do conjunto de dados. Em relação à metodologia aplicada, este experimento utilizou o método apresentado no primeiro experimento (fluxograma 13), uma vez que as imagens foram levadas ao sistema de OCR sem tratamento algum, apenas em escala de cinza. Entretanto,

Imagem	Resultado
Original, em escala de cinza	WOCONTEM (COMPOSIGAO (CARACTERGTERY GLUTEN. QUIMICA (mg/L) FISICO-UIMICAs Estrineo..... 0,037 CAICIO s, 25 : Magnésio..... 0,09 g POtS8I...vc... 0,21 [S0010....core. 75,81 "Vanédo.....,....mW Sulfato.....fi,28' e 81,10
Aplicado remoção de ruído	.u' - MOCONTEM COMPOSICAO (CARACTERGTE: GLUTEN. QUIMICA (mgL) (FisICo. : Estrinc..... 0,037 ' \\ CACIB it 1,25 \\ Magnésio..... 0,09 \\ = Potéssio..... 0,21 \\ '— 30di..... 15,81 \\ e T[T \\ ' OS .61, \\ :3' - Bicarbonato....
Aplicado abertura, remoção de ruído e Canny	iip CONTIER fi@@?fi@@fi?fi@f BLUTER QUIEMICA (mgL) %ffi@&%@%m@ - V= L il O fl @@afi&@z@@ @E?%%@@@WW@%@@ BT i CEeL g, YO0 e e A 7 [l e B R (010} S o — ol CGE Sl N N T B @ F=a IR TWTRE el ANIE N Ly S e ffi@ AR R U SNTRe . s 7 A e < // % <e

Tabela 4 – Resultado de uma imagem do segundo experimento, dividido pelas técnicas aplicadas

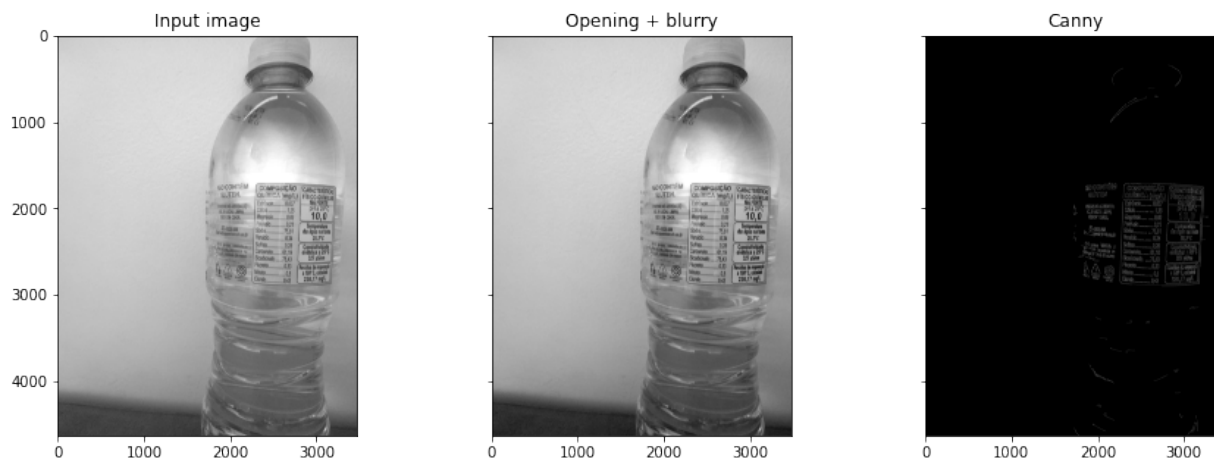


Figura 17 – Exemplo obtido no segundo grupo dos experimentos. À esquerda, a imagem original, em escala de cinza. No meio, a imagem aplicada à abertura e embaçamento. À direita, a imagem aplicada ao Canny.

pela natureza das imagens deste conjunto, foi possível obter resultados satisfatórios.

3.2.3.2 Resultados

A tabela 5 apresenta três exemplos de resultados obtidos. Para a primeira imagem, observa-se que foi possível obter o texto quase que integralmente, salvo algumas exceções de acentuação e caracteres especiais. A segunda imagem também apresentou um bom resultado; em geral, a maior parte das informações foi recuperada, com exceção da última linha. Já para a terceira imagem, verifica-se que a informação recuperada não foi tão boa. É possível notar que apenas a parte esquerda do rótulo foi previamente coberta, enquanto pouca coisa útil foi recuperada da direita.

3.2.4 Quarto experimento

3.2.4.1 Descrição

Para o quarto grupo de experimentos, foi implementado uma série de passos de pré-processamento, inspirados na técnica apresentada por (GONZALEZ *et al.*, 2012). Primeiramente, aplicou-se o algoritmo de MSER, a fim de obter regiões de interesse na imagem. Após, foi necessário realizar um filtro sobre as regiões obtidas, uma vez que a implementação utilizada gera um número muito alto, com várias repetições. Para a filtragem, as imagens foram comparadas par a par, sequencialmente, a partir do valor de intersecção sobre a união (IoU). Utilizou-se um limiar de 70% para decidir se as regiões eram iguais. A etapa de filtragem foi realmente necessária; em alguns casos, foi capaz de reduzir aproximadamente 80% as regiões obtidas no primeiro passo. Como o objetivo consiste na identificação de palavras, foi necessário realizar a junção das regiões, a fim de

Imagem	Resultado Obtido
	<p>CONCESSIONARIA: MINERADORA DE AGUAS RAINHA LTDA (NP):01.310.808/0001-38. INDUSTRIA BRASILEIRA. Distribuido e Comercializado por: Aimaleve Bebidas e Logistica - (NPJ: 18.119.353/0002-51 Avenida Brasil, 804 - Jardim Paran - CEP: 85935-000 Assis Chateaubriand - PR,FONTE BELA VISTA, LOCAL DA FONTE: ESTRADA DA MINA, KM 02, SÍTIO BELA VISTA, GLEBA VALENCIA, CEP: 86.750-000, IGUARACU - PR. PROCESSO Ne 826.201/1994 DNPM. PORTARIA DE LAVRA Ne 576 DE 18/12/1998, D.O.U DE 21/12/1998., BOLETIM Ne 073/LAMIN/CPRM/2018 DE 18/04/2018.</p>
	<p>Classificagio: Agua Mineral Fluoretada e Fracaimente Radioativa na Fonte. Andlise Quimica Boletim n° 212/LAMIN/18 de 25/07/18 LAMIN/CPRM, NAO CONTEM GLUTEN Composicao quimica (mg/L) BICARBONATO 2313 sODIO 2,707 CALCIO 2,849 MAGNESIO 2,350 CLORETO 2,150 SULFATO 0,690 POTASSIO 0,617 NITRATO 0,820 FLUORETO 0,040 0,017 BARIO,w “ II ”J 1 ¥ 9822 z</p>
	<p>NAO CONTEM GLUTEN. CONSERVAR AO ABRIGO DO SOL, EM LOCAL LIMPO, SECO E SEM ODOOR. Temperatura SAC-453125.5000 da 4gua na fonte Emil: sac@grupopereirabrasil com 25,2°C ? i Condutividade Data de envase, validade e . : h elétricaa 25°C Wmfig,“m.mvagd;”“dem i 327 pSlem</p> <p>Residuo de evaporagao 1 2180°C, calcul Css 220,17 mg/L</p>

Tabela 5 – Resultados obtidos no terceiro grupo de experimentos

aplicar, posteriormente, o motor de OCR. A junção das regiões deu-se, assim como na filtragem, sequencialmente, par a par. A cada passo, realizava-se uma leve expansão da região e verificava-se a intersecção horizontal entre elas; caso houvesse, eram unidas. Essa etapa de junção também ajudou a reduzir bastante o número de regiões. Por fim, para cada região obtida, realizou-se o seu recorte da imagem original e aplicou-se o sistema de OCR. O fluxograma deste experimento é apresentado na figura 18, enquanto a figura 19 apresenta um exemplo prático destas etapas.

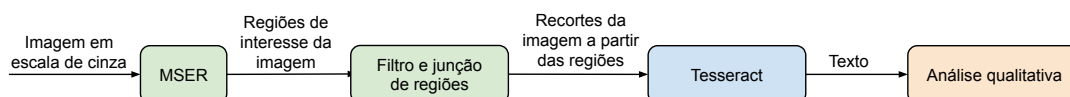


Figura 18 – Fluxograma do terceiro método, quarto experimento

3.2.4.2 Resultados

Um exemplo de resultado deste quarto grupo de experimentos é apresentado na tabela 6. Entretanto, observa-se que os resultados não foram tão satisfatórios, tanto na parte qualitativa como na parte de desempenho; aplicar o motor a cada região obtida foi um processo um tanto quanto demorado. Nota-se que algumas regiões necessárias, como na parte de composição química, foram removidas erroneamente. Ainda assim, ao contrário dos experimentos anteriores, foi possível capturar uma maior quantidade de regiões em uma imagem com bastante informação.

3.2.5 Quinto experimento

3.2.5.1 Descrição

O quinto grupo de experimentos foi realizado com o quarto conjunto (D) de dados. Nesse experimento, além do resgate das etapas de pré-processamento de imagem, foram adicionados passos de redimensionamento e de pós-processamento, permitindo uma análise quantitativa do resultado. O fluxograma 20 sumariza esse experimento.

Primeiramente, a imagem do rótulo é lida em escala de cinza. Após, a imagem redimensionada é levada a uma etapa de pré-processamento com técnicas de visão computacional clássica. Essa etapa de pré-processamento tem como objetivo realçar os caracteres do rótulo ao mesmo tempo que busca remover elementos da imagem que não sejam úteis para o reconhecimento ótico. Neste contexto, quatro técnicas são aplicadas à imagem.

A primeira técnica de tratamento de imagem é a suavização Gaussiana, para remoção de possíveis ruídos na imagem. Após, é aplicado limiarização adaptativa com o método de Otsu. Em seguida, para aumentar o contraste das regiões da imagem, é aplicado a técnica de equalização de histograma. Por fim, também com intuito de remover possíveis

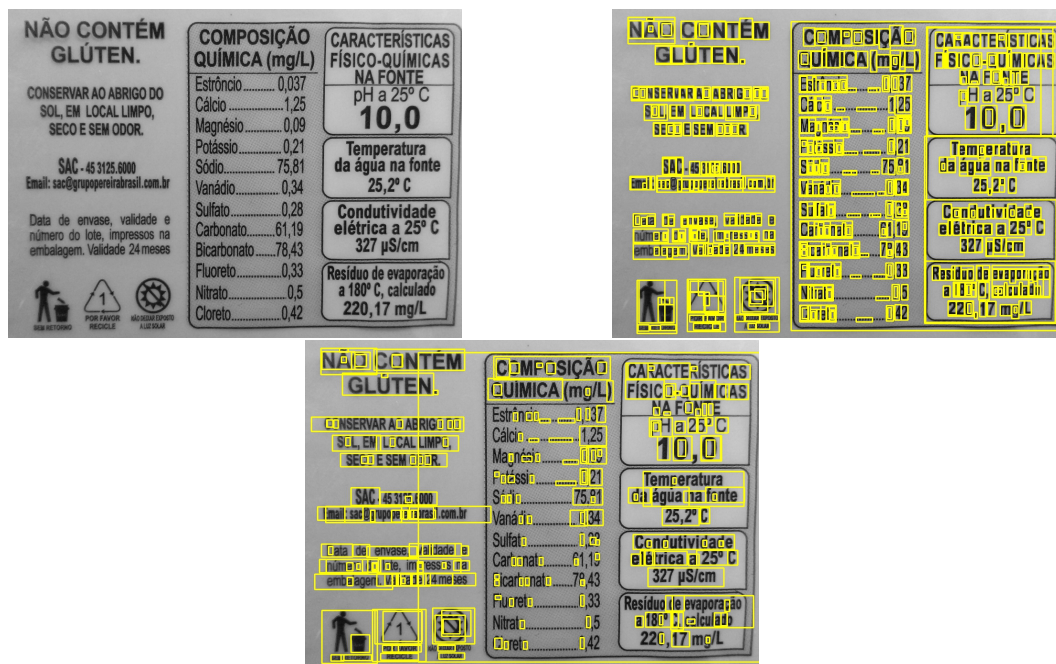


Figura 19 – Exemplo de processo do experimento quatro. Na esquerda superior, a imagem original em escala de cinza. Na direita superior, a imagem após MSER e filtragem de regiões. Na parte inferior centralizada, a imagem após junção das regiões.

ruídos, é aplicado a técnica de filtragem bilateral. Essa última técnica de suavização, em contraste com a primeira, tem seu principal diferencial no resguardo das bordas, fator importante para o reconhecimento dos caracteres.

Após as etapas de tratamento, a imagem é levada ao sistema de OCR. O texto obtido nesta etapa é aplicado a algumas etapas de processamento a fim de facilitar sua avaliação. Nesse processamento, todas as letras do texto são convertidas para minúsculas e os múltiplos espaços entre as palavras são removidos. Com isso, o texto obtido é convertido em uma lista de palavras.

A análise quantitativa do resultado é realizada avaliando a lista de palavras obtidas na etapa anterior com um gabarito de palavras-chave selecionadas. Nesse experimento, dois grupos de palavras-chave foram elencados. O primeiro considera os elementos necessários para rótulos de alimentos; o segundo, para rótulos de água. Assim, a distância de Levenshtein é computada entre cada palavra obtida para cada palavra-chave. Também é possível realizar a comparação de palavras compostas. Por fim, o resultado da análise consiste em uma relação numérica entre as palavras obtidas com as palavras-chave. Desse modo, busca-se, para cada palavra-chave, a palavra com o menor valor (menor distância); se esse valor estiver abaixo de um limiar pré-determinado, considera-se o termo como recuperado.

A acurácia, portanto, é computada a partir da razão entre a quantidade de termos recuperados e a quantidade de palavras-chave. São expostas duas acurácias: a maior

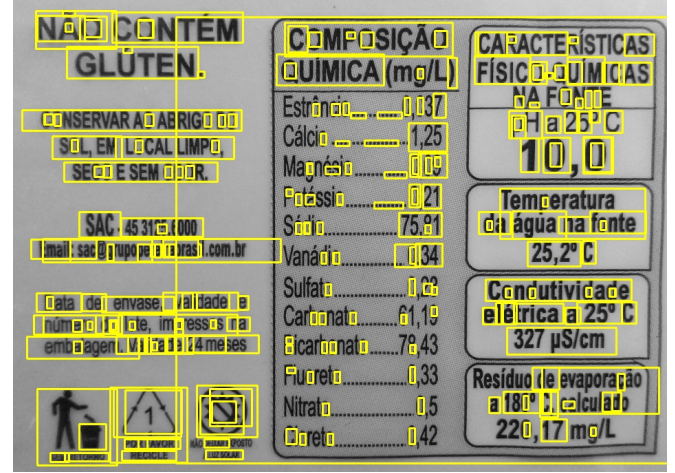
Imagem	Resultado Obtido
 <p>The image shows a water label for 'LOCAL LIMPO'. It includes the following information:</p> <ul style="list-style-type: none"> Gluten: NÃO CONTÉM GLÚTEN. Storage: CONSERVAR ABRIGADO, SELADO, EM LOCAL LIMPO, SECO E SEM ODORES. Contact: SAC: 45 3101 0000; Email: sac@grupopereira.com.br Shelf Life: Data de envase, validade e número lote impressos na embalagem. Validade de 24 meses. Chemical Composition (mg/L): <ul style="list-style-type: none"> Estrôncio: 0,137 Cálcio: 1,25 Magnésio: 0,09 Potássio: 0,21 Sódio: 75,81 Vanádio: 0,34 Sulfato: 0,02 Carbonato: 61,19 Bicarbonato: 78,43 Fluoreto: 0,33 Nitrato: 0,5 Cloreto: 0,42 Physical Characteristics: <ul style="list-style-type: none"> pH a 25°C: 10,0 Temperatura da água na fonte: 25,2°C Condutividade elétrica a 25°C: 327 µS/cm Resíduo de evaporação a 180°C calculado: 220,17 mg/L 	<p> LOCAL LIMPO, pereirabrasil.com.br sacOgrupoperei é > E TEM N., RIGO DO LIMPO, DOR.</p> <p>000 brasil.com.br validade e Dressos nNa 2 24meses &</p> <p>COMPOSIÇÃO (CARACTERÍSTICAS)) QUÍMICA (mg/L) FÍSICO QUÍMICAS POtássio.....m.... 0,21 Temperatura Sin 7581 da água na fonte Vanádio.....m.... 0,34 A ————— Condutividade elétrica a 25º C 327 uS/cm Resíduo de evapo a 180º C, calcul 220 17 mg/L COMPOSIÇÃO M o E 1,252/ pS/cm UIM Temperatura 3 água na fonte ic água </p>

Tabela 6 – Resultados obtidos no quarto grupo de experimentos

computada individualmente e a combinada, obtida através da união das listas de termos de todas as imagens processadas. Cada imagem foi avaliada com o seu respectivo grupo de palavras-chave.

3.2.5.2 Resultados

O resultado desse grupo de experimentos é apresentando nas tabelas 7 e 8.

A tabela 7 apresenta algumas imagens nas quais foram aplicadas as técnicas do experimento, assim como a sua respectiva imagem original. A coluna de identificação serve

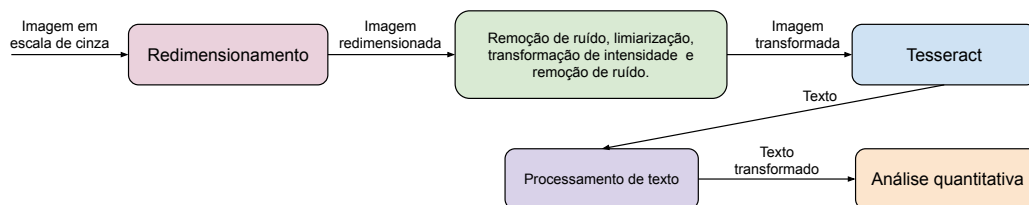


Figura 20 – Fluxograma do quarto método, quinto experimento

para guiar a visualização de resultados na tabela 8.

A tabela 8 apresenta os resultados das métricas computadas para as imagens da tabela 7. Como citado anteriormente, a maior acurácia individual, entre os processamentos realizados para cada fator de redimensionamento, é exposta. Nota-se, para os identificadores 1 e 4 uma diferença significativa em relação à acurácia combinada, que leva em consideração os termos obtidos em todos os processamentos da mesma imagem. Por exemplo, para o identificador 1, apenas "gorduras saturadas"¹ não foi recuperado nesse processamento individual em questão, mas, sim, em algum outro. Já no caso dos identificadores 2 e 3, o melhor resultado individual já foi capaz de recuperar todas as palavras-chave.

Dois fatores principais diferem esse grupo de experimentos com o restante. Primeiramente, nota-se, qualitativamente, o realce nos caracteres dos rótulos, promovido pela aplicação das técnicas de pré-processamento. Isso fez com que o texto obtido no OCR fosse mais limpo que anteriormente. Por fim, a análise quantitativa, a partir da utilização da distância de Levensthein, permitiu uma observação prática dos resultados.

Ainda com os avanços realizados nesse grupo de experimento, não foi possível obter, de forma consistente, os valores numéricos de cada palavra-chave do rótulo. A tabela 9 apresenta um exemplo² desse fato. Nota-se, no texto de saída obtido para a primeira imagem apresentada, a falta dos valores numéricos de cada linha dos elementos. Alguns casos, como dos termos de "Valor Energético" e "Proteínas", tiveram uma recuperação de dígitos, mas de uma forma ruidosa. Entretanto, a parte inferior do rótulo, que trata da elucidação da escala de valores diários, apresentou um texto de saída preciso, inclusive com os números e unidade de medida em questão: 2000 kcal e 8400 kJ. Entretanto, para a segunda imagem, o cenário é outro. É possível verificar que a maior parte dos dígitos foram recuperados integralmente. Isso possibilita a formação de pares entre as palavras-chave com os seus respectivos valores. O conjunto combinado e de melhor resultado dessa recuperação é apresentado na tabela 10. Essa distinção de resultados acontece por algumas limitações do Tesseract, que necessita do contexto dos caracteres para um melhor reconhecimento.

¹ Palavras compostas são consideradas como apenas um termo

² Algumas quebras de linha foram adicionadas ao texto para facilitar a visualização

3.3 SUMÁRIO QUANTITATIVO DOS RESULTADOS

A etapa de quantificação dos resultados só foi implementada a partir do quinto grupo de experimentos. Desse modo, com intuito de apresentar métricas para os experimentos anteriores e possibilitar uma comparação formal, realizou-se mais uma etapa de experimentos. Para isso, foram elencadas cinco amostras de imagens dos conjuntos de dados A, C e D. Cada amostra de dados foi aplicada aos métodos já apresentados e, posteriormente, à etapa de processamento de texto e recuperação da métrica de acurácia. A tabela 11 apresenta os resultados obtidos, onde são elencadas a acurácia mínima, máxima e média para cada aplicação de amostra a um método. O maior resultado de cada grupo está em negrito, enquanto o segundo maior, sublinhado.

A coluna **Amostra** da tabela 11 identifica de qual conjunto de dados ela foi originada. A coluna **Método (descrição)** identifica qual foi o método aplicado às imagens daquela amostra, fazendo uma conexão com o diagrama de apresentação do método. Um método número 5 foi adicionado apenas para essa etapa; sendo considerado como uma ramificação do método 4. Nesse método, apenas o redimensionamento foi considerado na etapa de pré-processamento da imagem. As colunas seguintes apresentam as estatísticas referentes à **Acurácia** obtida em cada aplicação.

Os resultados dessa seção reforçam o que já foi apresentado anteriormente, principalmente em relação às características da imagem. É notável que os melhores resultados foram obtidos com amostras do conjunto de dados D, enquanto os piores resultados foram obtidos com a amostra do primeiro conjunto. Além disso, notamos que o último experimento (número 5) apresentou um resultado melhor, comparado com os anteriores. De fato, os melhores resultados foram com os métodos 4 e 5. Ainda assim, um dos pontos principais e de atenção desse experimento é, de fato, o redimensionamento. Isso se dá pois o método 5 foi capaz de demonstrar acurácia média maior em todos os casos, ainda que os resultados sejam bem próximos.

3.4 APLICAÇÃO

Um protótipo de aplicação foi desenvolvido para utilização interativa do fluxo de execução proposto no experimento 3.2.5. O aplicativo **paNNe**³ foi desenvolvido na linguagem Python e utilizando o framework Streamlit para possibilitar a visualização e interação do usuário. Sua exibição está disponível tanto em português como em inglês.

A figura 21 apresenta o cabeçalho de apresentação do aplicativo, com seu título, subtítulo e o logo da Universidade Federal de Santa Catarina.

As imagens na figura 22 apresentam as definições disponíveis para processamento de imagem e de texto. Na imagem da esquerda estão apresentadas as possíveis customizações no fluxo de pré-processamento que a imagem passará. É possível selecionar quais

³ O código fonte está disponível em <https://github.com/cchenzi/tcc>.



paNNe

Package Label Scanner

Figura 21 – Cabeçalho do aplicativo

métodos serão utilizados, assim como quais fatores de escala devem ser aplicados no redimensionamento. A imagem da direita apresenta a definição do pós-processamento e contabilização de métricas do texto de saída do sistema. O usuário seleciona um conjunto de palavras que será utilizado na avaliação do resultado; é possível, também, remover palavras pré-definidas.

A figura 23 demonstra como são apresentadas as imagens após serem aplicadas as técnicas de processamento. Uma imagem resultante é exposta para cada fator de escala aplicado. Com intuito de permitir uma melhor navegação e economizar espaço na tela, sua exibição só ocorre, de fato, após o usuário clicar no respectivo *drop-down*.

A seção de resultados é apresentada logo após a finalização da avaliação de cada imagem individualmente. Desse modo, são apresentadas avaliações combinadas, onde o resultado de todas as imagens avaliadas são considerados, e a melhor avaliação individual. Essa seção é apresentada na figura 24. Além da acurácia obtida, também são apresentadas as listas de termos recuperados, o dicionário de dígitos recuperados e o texto recuperado para a imagem com melhor avaliação.

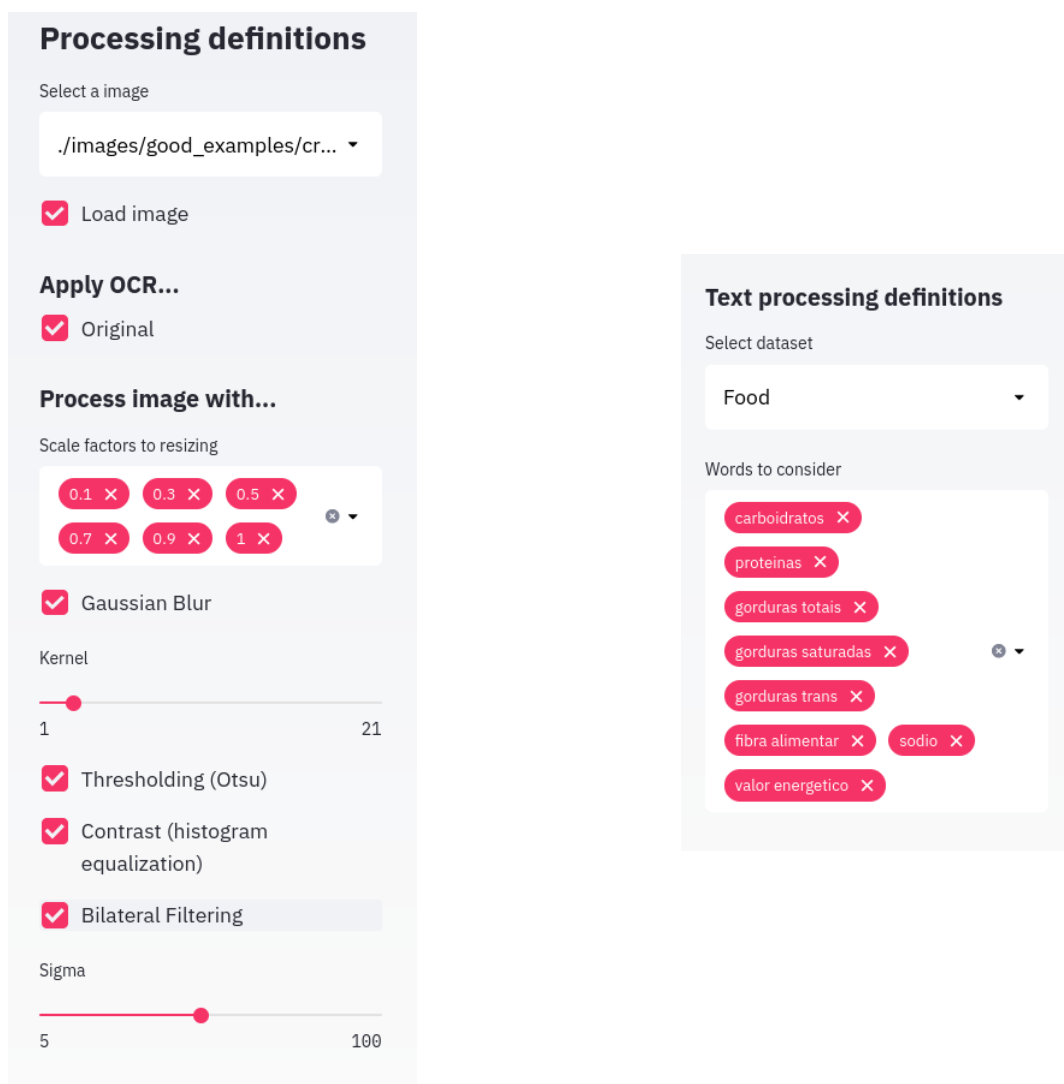


Figura 22 – Definições de processamento de imagem e texto

Identificador	Original	Processada																																																												
1	<p>INFORMAÇÃO NUTRICIONAL Porção de 15 g (1 colher e ½ de sopa)</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th></th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor energético</td> <td>26 kcal = 109 kJ</td> <td>1</td> </tr> <tr> <td>Carboidratos</td> <td>0,6 g</td> <td>0</td> </tr> <tr> <td>Proteínas</td> <td>0</td> <td>0</td> </tr> <tr> <td>Gorduras totais</td> <td>2,6 g</td> <td>5</td> </tr> <tr> <td>Gorduras saturadas</td> <td>1,6 g</td> <td>7</td> </tr> <tr> <td>Gorduras trans</td> <td>0</td> <td>**</td> </tr> <tr> <td>Fibra alimentar</td> <td>0</td> <td>0</td> </tr> <tr> <td>Sódio</td> <td>12 mg</td> <td>1</td> </tr> </tbody> </table> <p>* % Valores Diários com base em uma dieta de 2.000 kcal ou 8.400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas. ** VD não estabelecido.</p>	Quantidade por porção		% VD (*)	Valor energético	26 kcal = 109 kJ	1	Carboidratos	0,6 g	0	Proteínas	0	0	Gorduras totais	2,6 g	5	Gorduras saturadas	1,6 g	7	Gorduras trans	0	**	Fibra alimentar	0	0	Sódio	12 mg	1	<p>▼ 0.5</p> <p>INFORMAÇÃO NUTRICIONAL Porção de 15 g (1 colher e ½ de sopa)</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th></th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor energético</td> <td>26 kcal = 109 kJ</td> <td>1</td> </tr> <tr> <td>Carboidratos</td> <td>0,6 g</td> <td>0</td> </tr> <tr> <td>Proteínas</td> <td>0</td> <td>0</td> </tr> <tr> <td>Gorduras totais</td> <td>2,6 g</td> <td>5</td> </tr> <tr> <td>Gorduras saturadas</td> <td>1,6 g</td> <td>7</td> </tr> <tr> <td>Gorduras trans</td> <td>0</td> <td>**</td> </tr> <tr> <td>Fibra alimentar</td> <td>0</td> <td>0</td> </tr> <tr> <td>Sódio</td> <td>12 mg</td> <td>1</td> </tr> </tbody> </table> <p>* % Valores Diários com base em uma dieta de 2.000 kcal ou 8.400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas. ** VD não estabelecido.</p>	Quantidade por porção		% VD (*)	Valor energético	26 kcal = 109 kJ	1	Carboidratos	0,6 g	0	Proteínas	0	0	Gorduras totais	2,6 g	5	Gorduras saturadas	1,6 g	7	Gorduras trans	0	**	Fibra alimentar	0	0	Sódio	12 mg	1						
Quantidade por porção		% VD (*)																																																												
Valor energético	26 kcal = 109 kJ	1																																																												
Carboidratos	0,6 g	0																																																												
Proteínas	0	0																																																												
Gorduras totais	2,6 g	5																																																												
Gorduras saturadas	1,6 g	7																																																												
Gorduras trans	0	**																																																												
Fibra alimentar	0	0																																																												
Sódio	12 mg	1																																																												
Quantidade por porção		% VD (*)																																																												
Valor energético	26 kcal = 109 kJ	1																																																												
Carboidratos	0,6 g	0																																																												
Proteínas	0	0																																																												
Gorduras totais	2,6 g	5																																																												
Gorduras saturadas	1,6 g	7																																																												
Gorduras trans	0	**																																																												
Fibra alimentar	0	0																																																												
Sódio	12 mg	1																																																												
2	<p>Classificação: Água Mineral Fluoretada e Fracamente Radioativa na Fonte. Análise Química Boletim nº 212/LAMIN/18 de 25/07/18 LAMIN/CPRM. NÃO CONTEM GLUTEN</p> <table border="1"> <thead> <tr> <th>Composição química (mg/L)</th> <th></th> </tr> </thead> <tbody> <tr> <td>BICARBONATO</td> <td>23,13</td> </tr> <tr> <td>SÓDIO</td> <td>2,707</td> </tr> <tr> <td>CÁLCIO</td> <td>2,849</td> </tr> <tr> <td>MAGNÉSIO</td> <td>2,350</td> </tr> <tr> <td>CLORETO</td> <td>2,150</td> </tr> <tr> <td>SULFATO</td> <td>0,690</td> </tr> <tr> <td>POTÁSSIO</td> <td>0,617</td> </tr> <tr> <td>NITRATO</td> <td>0,820</td> </tr> <tr> <td>FLUORETO</td> <td>0,040</td> </tr> <tr> <td>ESTRÔNCIO</td> <td>0,017</td> </tr> <tr> <td>BÁRIO</td> <td>0,014</td> </tr> </tbody> </table>	Composição química (mg/L)		BICARBONATO	23,13	SÓDIO	2,707	CÁLCIO	2,849	MAGNÉSIO	2,350	CLORETO	2,150	SULFATO	0,690	POTÁSSIO	0,617	NITRATO	0,820	FLUORETO	0,040	ESTRÔNCIO	0,017	BÁRIO	0,014	<p>▼ 0.5</p> <p>Classificação: Água Mineral Fluoretada e Fracamente Radioativa na Fonte. Análise Química Boletim nº 212/LAMIN/18 de 25/07/18 LAMIN/CPRM. NÃO CONTEM GLUTEN</p> <table border="1"> <thead> <tr> <th>Composição química (mg/L)</th> <th></th> </tr> </thead> <tbody> <tr> <td>BICARBONATO</td> <td>23,13</td> </tr> <tr> <td>SÓDIO</td> <td>2,707</td> </tr> <tr> <td>CÁLCIO</td> <td>2,849</td> </tr> <tr> <td>MAGNÉSIO</td> <td>2,350</td> </tr> <tr> <td>CLORETO</td> <td>2,150</td> </tr> <tr> <td>SULFATO</td> <td>0,690</td> </tr> <tr> <td>POTÁSSIO</td> <td>0,617</td> </tr> <tr> <td>NITRATO</td> <td>0,820</td> </tr> <tr> <td>FLUORETO</td> <td>0,040</td> </tr> <tr> <td>ESTRÔNCIO</td> <td>0,017</td> </tr> <tr> <td>BÁRIO</td> <td>0,014</td> </tr> </tbody> </table>	Composição química (mg/L)		BICARBONATO	23,13	SÓDIO	2,707	CÁLCIO	2,849	MAGNÉSIO	2,350	CLORETO	2,150	SULFATO	0,690	POTÁSSIO	0,617	NITRATO	0,820	FLUORETO	0,040	ESTRÔNCIO	0,017	BÁRIO	0,014												
Composição química (mg/L)																																																														
BICARBONATO	23,13																																																													
SÓDIO	2,707																																																													
CÁLCIO	2,849																																																													
MAGNÉSIO	2,350																																																													
CLORETO	2,150																																																													
SULFATO	0,690																																																													
POTÁSSIO	0,617																																																													
NITRATO	0,820																																																													
FLUORETO	0,040																																																													
ESTRÔNCIO	0,017																																																													
BÁRIO	0,014																																																													
Composição química (mg/L)																																																														
BICARBONATO	23,13																																																													
SÓDIO	2,707																																																													
CÁLCIO	2,849																																																													
MAGNÉSIO	2,350																																																													
CLORETO	2,150																																																													
SULFATO	0,690																																																													
POTÁSSIO	0,617																																																													
NITRATO	0,820																																																													
FLUORETO	0,040																																																													
ESTRÔNCIO	0,017																																																													
BÁRIO	0,014																																																													
3	<p>INGREDIENTES: leite integral, estabilizantes: polifosfato sódico (E401), E402, citrato de sódio (E301), emulsificantes: mono e diglicerol de ácidos graxos (E471), E472.</p> <p>ALERGICOS: CONTEM LEITE. CONTEM LACTOSE. NÃO CONTEM GLUTEN.</p> <p>INFORMAÇÃO NUTRICIONAL Porção de 200 ml (1 copo)</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th></th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor Energético</td> <td>116 kcal = 487 kJ</td> <td>6</td> </tr> <tr> <td>Carboidratos</td> <td>9,1 g</td> <td>3</td> </tr> <tr> <td>Proteínas</td> <td>4,3 g</td> <td>8</td> </tr> <tr> <td>Gorduras Totais</td> <td>6,0 g</td> <td>11</td> </tr> <tr> <td>Gorduras Saturadas</td> <td>4,3 g</td> <td>20</td> </tr> <tr> <td>Gorduras Trans</td> <td>0 g</td> <td>**</td> </tr> <tr> <td>Fibra Alimentar</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Sódio</td> <td>152 mg</td> <td>6</td> </tr> <tr> <td>Cálcio</td> <td>209 mg</td> <td>21</td> </tr> </tbody> </table> <p>* % Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** Valor não especificado.</p>	Quantidade por porção		% VD (*)	Valor Energético	116 kcal = 487 kJ	6	Carboidratos	9,1 g	3	Proteínas	4,3 g	8	Gorduras Totais	6,0 g	11	Gorduras Saturadas	4,3 g	20	Gorduras Trans	0 g	**	Fibra Alimentar	0 g	0	Sódio	152 mg	6	Cálcio	209 mg	21	<p>▼ 0.9</p> <p>INFORMAÇÃO NUTRICIONAL Porção de 200 ml (1 copo)</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th></th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor Energético</td> <td>116 kcal = 487 kJ</td> <td>6</td> </tr> <tr> <td>Carboidratos</td> <td>9,1 g</td> <td>3</td> </tr> <tr> <td>Proteínas</td> <td>6,3 g</td> <td>8</td> </tr> <tr> <td>Gorduras Totais</td> <td>6,0 g</td> <td>11</td> </tr> <tr> <td>Gorduras Saturadas</td> <td>4,3 g</td> <td>20</td> </tr> <tr> <td>Gorduras Trans</td> <td>0 g</td> <td>**</td> </tr> <tr> <td>Fibra Alimentar</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Sódio</td> <td>152 mg</td> <td>6</td> </tr> <tr> <td>Cálcio</td> <td>209 mg</td> <td>21</td> </tr> </tbody> </table> <p>* % Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** Valor não especificado.</p>	Quantidade por porção		% VD (*)	Valor Energético	116 kcal = 487 kJ	6	Carboidratos	9,1 g	3	Proteínas	6,3 g	8	Gorduras Totais	6,0 g	11	Gorduras Saturadas	4,3 g	20	Gorduras Trans	0 g	**	Fibra Alimentar	0 g	0	Sódio	152 mg	6	Cálcio	209 mg	21
Quantidade por porção		% VD (*)																																																												
Valor Energético	116 kcal = 487 kJ	6																																																												
Carboidratos	9,1 g	3																																																												
Proteínas	4,3 g	8																																																												
Gorduras Totais	6,0 g	11																																																												
Gorduras Saturadas	4,3 g	20																																																												
Gorduras Trans	0 g	**																																																												
Fibra Alimentar	0 g	0																																																												
Sódio	152 mg	6																																																												
Cálcio	209 mg	21																																																												
Quantidade por porção		% VD (*)																																																												
Valor Energético	116 kcal = 487 kJ	6																																																												
Carboidratos	9,1 g	3																																																												
Proteínas	6,3 g	8																																																												
Gorduras Totais	6,0 g	11																																																												
Gorduras Saturadas	4,3 g	20																																																												
Gorduras Trans	0 g	**																																																												
Fibra Alimentar	0 g	0																																																												
Sódio	152 mg	6																																																												
Cálcio	209 mg	21																																																												
4	<p>TABELA NUTRICIONAL Porção de 100 g (1 porção)</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th></th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor Energético</td> <td>3Kcal / 12kJ</td> <td>0</td> </tr> <tr> <td>Carboidratos</td> <td>2,3 g</td> <td>1</td> </tr> <tr> <td>Proteínas</td> <td>0,3 g</td> <td>0</td> </tr> <tr> <td>Gorduras Totais</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Gorduras Saturadas</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Gorduras Trans</td> <td>0 g</td> <td>**</td> </tr> <tr> <td>Fibra Alimentar</td> <td>1,9 g</td> <td>7</td> </tr> <tr> <td>Sódio</td> <td>0,3 mg</td> <td>0</td> </tr> </tbody> </table>	Quantidade por porção		% VD (*)	Valor Energético	3Kcal / 12kJ	0	Carboidratos	2,3 g	1	Proteínas	0,3 g	0	Gorduras Totais	0 g	0	Gorduras Saturadas	0 g	0	Gorduras Trans	0 g	**	Fibra Alimentar	1,9 g	7	Sódio	0,3 mg	0	<p>▼ 0.3</p> <p>TABELA NUTRICIONAL Porção de 100 g (1 porção)</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th></th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor Energético</td> <td>3Kcal / 12kJ</td> <td>0</td> </tr> <tr> <td>Carboidratos</td> <td>2,3 g</td> <td>1</td> </tr> <tr> <td>Proteínas</td> <td>0,3 g</td> <td>0</td> </tr> <tr> <td>Gorduras Totais</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Gorduras Saturadas</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Gorduras Trans</td> <td>0 g</td> <td>**</td> </tr> <tr> <td>Fibra Alimentar</td> <td>1,9 g</td> <td>7</td> </tr> <tr> <td>Sódio</td> <td>0,3 mg</td> <td>0</td> </tr> </tbody> </table>	Quantidade por porção		% VD (*)	Valor Energético	3Kcal / 12kJ	0	Carboidratos	2,3 g	1	Proteínas	0,3 g	0	Gorduras Totais	0 g	0	Gorduras Saturadas	0 g	0	Gorduras Trans	0 g	**	Fibra Alimentar	1,9 g	7	Sódio	0,3 mg	0						
Quantidade por porção		% VD (*)																																																												
Valor Energético	3Kcal / 12kJ	0																																																												
Carboidratos	2,3 g	1																																																												
Proteínas	0,3 g	0																																																												
Gorduras Totais	0 g	0																																																												
Gorduras Saturadas	0 g	0																																																												
Gorduras Trans	0 g	**																																																												
Fibra Alimentar	1,9 g	7																																																												
Sódio	0,3 mg	0																																																												
Quantidade por porção		% VD (*)																																																												
Valor Energético	3Kcal / 12kJ	0																																																												
Carboidratos	2,3 g	1																																																												
Proteínas	0,3 g	0																																																												
Gorduras Totais	0 g	0																																																												
Gorduras Saturadas	0 g	0																																																												
Gorduras Trans	0 g	**																																																												
Fibra Alimentar	1,9 g	7																																																												
Sódio	0,3 mg	0																																																												

Tabela 7 – Resultados obtidos no quinto grupo de experimentos

Identificador	Melhor Acurácia Individual	Termos recuperados (individual)	Acurácia Combinada	Termos recuperados (combinado)
1	0.87	"carboidratos", "proteinas", "gorduras totais", "gorduras trans", "fibra alimentar", "sodio", "valor energetico"	1	"valor energetico", "carboidratos", "gorduras saturadas", "proteinas", "gorduras totais", "fibra alimentar", "sodio", "gorduras trans"
2	1	"fluoreto", "bario", "sodio", "sulfato", "cloreto", "magnésio", "bicarbonato", "nitrate", "calcio", "estroncio", "potassio"	1	"fluoreto", "bario", "sodio", "sulfato", "cloreto", "magnésio", "bicarbonato", "nitrate", "calcio", "estroncio", "potassio"
3	1	"valor energetico", "carboidratos", "gorduras saturadas", "proteinas", "gorduras totais", "fibra alimentar", "sodio", "gorduras trans"	1	"valor energetico", "carboidratos", "gorduras saturadas", "proteinas", "gorduras totais", "fibra alimentar", "sodio", "gorduras trans"
4	0.5	"proteinas", "gorduras totais", "gorduras saturadas", "gorduras trans"	0.62	"gorduras saturadas", "proteinas", "gorduras totais", "fibra alimentar", "gorduras trans"

Tabela 8 – Resultados obtidos no quinto grupo de experimentos

Imagem	Resultado Obtido																														
<p>▼ 0.9</p> <table border="1"> <caption>INFORMAÇÃO NUTRICIONAL Porção de 200 ml (1 copo)</caption> <thead> <tr> <th></th> <th>Quantidade por porção</th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor Energético</td> <td>116 kcal = 487 kJ</td> <td>6</td> </tr> <tr> <td>Carboidratos</td> <td>9,1 g</td> <td>3</td> </tr> <tr> <td>Proteínas</td> <td>6,3 g</td> <td>8</td> </tr> <tr> <td>Gorduras Totais</td> <td>6,0 g</td> <td>11</td> </tr> <tr> <td>Gorduras Saturadas</td> <td>4,3 g</td> <td>20</td> </tr> <tr> <td>Gorduras Trans</td> <td>0 g</td> <td>**</td> </tr> <tr> <td>Fibra Alimentar</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Sódio</td> <td>152 mg</td> <td>6</td> </tr> <tr> <td>Cálcio</td> <td>209 mg</td> <td>21</td> </tr> </tbody> </table> <p>* % Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** Valor não especificado.</p>		Quantidade por porção	% VD (*)	Valor Energético	116 kcal = 487 kJ	6	Carboidratos	9,1 g	3	Proteínas	6,3 g	8	Gorduras Totais	6,0 g	11	Gorduras Saturadas	4,3 g	20	Gorduras Trans	0 g	**	Fibra Alimentar	0 g	0	Sódio	152 mg	6	Cálcio	209 mg	21	<p>INFORMAÇÃO NUTRICIONAL</p> <p>Porção de 200 ml (1 copo) Quantidade por porção Valor Energético — 116kcal=487kJ 6 8 Carboidratos Proteínas 63g g</p> <p>Gorduras Totais</p> <p>Gorduras Saturadas</p> <p>Gorduras Trans Fibra Alimentar</p> <p>ódio Cálcio</p> <p>% Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** Valor não especificado.</p>
	Quantidade por porção	% VD (*)																													
Valor Energético	116 kcal = 487 kJ	6																													
Carboidratos	9,1 g	3																													
Proteínas	6,3 g	8																													
Gorduras Totais	6,0 g	11																													
Gorduras Saturadas	4,3 g	20																													
Gorduras Trans	0 g	**																													
Fibra Alimentar	0 g	0																													
Sódio	152 mg	6																													
Cálcio	209 mg	21																													
<p>▼ 0.5</p> <p>Classificação: Água Mineral Fluoretada e Fracamente Radioativa na Fonte. Análise Química Boletim nº 212/LAMIN/18 de 25/07/18 LAMIN/CPRM. NÃO CONTÉM GLUTÊN</p> <p>Composição química (mg L)</p> <table border="1"> <tbody> <tr> <td>BICARBONATO</td> <td>23,13</td> </tr> <tr> <td>SÓDIO</td> <td>2,707</td> </tr> <tr> <td>CÁLCIO</td> <td>2,849</td> </tr> <tr> <td>MAGNÉSIO</td> <td>2,350</td> </tr> <tr> <td>CLORETO</td> <td>2,150</td> </tr> <tr> <td>SULFATO</td> <td>0,690</td> </tr> <tr> <td>POTÁSSIO</td> <td>0,617</td> </tr> <tr> <td>NITRATO</td> <td>0,820</td> </tr> <tr> <td>FLUORETO</td> <td>0,648</td> </tr> <tr> <td>EXTRÔNCIO</td> <td>0,017</td> </tr> <tr> <td>DÁRIO</td> <td>0,014</td> </tr> </tbody> </table>	BICARBONATO	23,13	SÓDIO	2,707	CÁLCIO	2,849	MAGNÉSIO	2,350	CLORETO	2,150	SULFATO	0,690	POTÁSSIO	0,617	NITRATO	0,820	FLUORETO	0,648	EXTRÔNCIO	0,017	DÁRIO	0,014	<p>Classificação: Água Mineral Fluoretada é Fracamente Radioativa na Fonte. Análise Química Boletim nº 212/ L AMIN/18 de 25/07/18 LAMIN/CPRM. NÃO CONTÉM GLUTEN BICARBONATO 23,13 SÓDIO 2,707 CÁLCIO 2,649 MAGNÉSIO 2,350 CLORETO 2,150 SULFATO 0,690 POTÁSSIO 0617 MITRATO 0,820 FLUORETO 0,648 EXTRÔNCIO e DÁRIO 01</p>								
BICARBONATO	23,13																														
SÓDIO	2,707																														
CÁLCIO	2,849																														
MAGNÉSIO	2,350																														
CLORETO	2,150																														
SULFATO	0,690																														
POTÁSSIO	0,617																														
NITRATO	0,820																														
FLUORETO	0,648																														
EXTRÔNCIO	0,017																														
DÁRIO	0,014																														

Tabela 9 – Resultados obtido no quinto grupo de experimentos

Identificador	Melhor Resultado	Resultado Combinado
2	"bicarbonato": "23,13", "sodio": "2,707", "calcio": "2,649", "magnesio": "2,350", "cloreto": "2,150", "sulfato": "0,690", "potassio": "0617", "nittrato": "0,820", "fluoreto": "0,648", "bario": "01"	"bicarbonato": ["23,13"], "sodio": ["2,707"], "calcio": ["2,849", "2,649"], "magnesio": ["2,350"], "cloreto": ["2,150"], "sulfato": ["0,690"], "potassio": ["0617", "0,617"], "nittrato": ["0,820"], "fluoreto": ["0,648", "0,640"], "bario": ["0,014", "01"]

Tabela 10 – Exemplo de recuperação de dígitos no quinto grupo de experimentos

Amostra	Método (descrição)	Acurácia Mínima	Acurácia Máxima	Acurácia Média
1	1 (fig. 13)	0	0.27	0.12
1	2 (fig. 16)	0	0	0
1	4 (fig. 20)	0	<u>0.54</u>	<u>0.25</u>
1	5 (fig. 20) *	0	0.63	0.29
3	1 (fig. 13)	0	0.27	0.07
3	2 (fig. 16)	0	<u>0.36</u>	0.07
3	4 (fig. 20)	0.09	1	<u>0.52</u>
3	5 (fig. 20) *	0.09	1	0.56
4	1 (fig. 13)	0	1	0.3
4	2 (fig. 16)	0	<u>0.75</u>	0.29
4	4 (fig. 20)	<u>0.81</u>	1	<u>0.93</u>
4	5 (fig. 20) *	0.87	1	0.95

Tabela 11 – Sumário quantitativo para comparação dos métodos apresentados.

* A etapa de transformação da imagem não foi realizada.

Em negrito, destaca-se a melhor acurácia para aquele grupo; sublinhada, a segunda melhor acurácia.

paNNe

Package Label Scanner

- ▶ Original
-
- ▶ 0.1
-
- ▶ 0.3
-
- ▶ 0.5
-
- ▶ 0.7
-
- ▶ 0.9
-
- ▶ 1

paNNe

Package Label Scanner

- ▶ Original
-
- ▶ 0.1
-
- ▶ 0.3
-
- ▼ 0.5

INFORMAÇÃO NUTRICIONAL
Porção de 15 g (1 colher e ½ de sopa)

Quantidade por porção		% VD (*)
Valor energético	26 kcal = 109 kJ	1
Carboidratos	0,6 g	0
Proteínas	0	0
Gorduras totais	2,6 g	5
Gorduras saturadas	1,6 g	7
Gorduras trans	0	**
Fibra alimentar	0	0
Sódio	12 mg	1

* % Valores Diários com base em uma dieta de 2.000 kcal ou 8.400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas.
 ** VD não estabelecido.

Figura 23 – Exibição das imagens processadas

Accuracy

Combined result acc = 1.0

Best result for scale_factor = 0.5, with acc = 0.875!

Terms recovered (combined)

```
▼ [
  0 : "valor energetico"
  1 : "carboidratos"
  2 : "gorduras saturadas"
  3 : "proteinas"
  4 : "gorduras totais"
  5 : "fibra alimentar"
  6 : "sodio"
  7 : "gorduras trans"
]
```

Terms recovered (best result)

```
▼ [
  0 : "carboidratos"
  1 : "proteinas"
  2 : "gorduras totais"
  3 : "gorduras trans"
  4 : "fibra alimentar"
  5 : "sodio"
  6 : "valor energetico"
]
```

Recovered text...

INFORMAÇÃO NUTRICIONAL Porção de 15 g (1 colher e %4 de sopa) Quantidade por porção S
VDf) Valor energético ; Carboidratos : Proteinas

Gorduras totais

Gorduras trans Fibra alimentar

Sódio *% Valores Diários com base em uma dieta de 2.000 kcal ou 8.400 kJ, Seus valores
diários podem ser maiores ou necessidades energéticas.

Figura 24 – Apresentação dos resultados

4 CONCLUSÃO

O reconhecimento óptico de caracteres, ainda que amplamente estudado na literatura, ainda apresenta desafios a serem ultrapassados. Diversas técnicas assumem características específicas para apresentar uma boa performance. Nesse sentido, etapas de pré-processamento de imagens são essenciais para obter resultados que sejam úteis no cotidiano. Neste trabalho, foi apresentado uma visão geral sobre a leitura automática de rótulos de embalagens, com enfoque na utilização de um sistema de OCR de código aberto, o Tesseract. Foram realizados experimentos exploratórios, de avaliação qualitativa e quantitativa. Em ambos os cenários metodologias de pré-processamento de imagens foram abordadas e utilizadas.

Os experimentos exploratórios de avaliação qualitativa tiveram intuito de promover o entendimento sobre o comportamento do Tesseract e ampliar o conhecimento sobre técnicas de visão computacional clássica. Por se tratar de análise qualitativa, a discussão dos resultados é um tanto quanto abstrata. Notou-se que vários textos foram recuperados com termos úteis, mas que, ainda assim, deixavam a desejar em questão de consistência. Entretanto, esses experimentos foram úteis para pavimentar a trajetória para as análises quantitativas, com definições e objetivos mais concretos.

Os experimentos de avaliação quantitativa mostraram bons resultados com imagens em condições favoráveis (e.g., foco no rótulo, boa iluminação, rótulo planificado) . Foi possível recuperar palavras-chave dos rótulos com uma grande precisão. Isso permite uma análise elementar dos fatos nutricionais presentes nas embalagens; torna-se, portanto, capaz de verificar as conformidades de um rótulo em relação à presença de informações básicas. Também foi apresentado um exemplo da produção de pares entre palavras-chave e seus respectivos valores numéricos. Entretanto, essa última parte ainda é muito sensível às características da imagem, uma vez que não foi possível recuperar, de forma consistente, para o restante das imagens. Em razão disso, a utilização do protótipo apresentado pode ser um pouco restrita.

Para continuação deste projeto, trabalhos futuros incluem melhorar as técnicas de pré-processamento de imagens já apresentadas, assim como explorar e desenvolver novos algoritmos que contribuam neste aspecto. Uma das possíveis melhorias consiste em realizar a extração consistente de regiões, a fim de diminuir o preparo manual que antecede à aplicação das técnicas. Além disso, também é factível o aperfeiçoamento na recuperação de caracteres numéricos, com intuito de permitir análises mais profundas em relação às necessidades e informações dos rótulos de embalagens.

REFERÊNCIAS

- AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA - ANVISA. **Rotulagem de alimentos**. [S.l.: s.n.], 2020. [Online; acessado em 08/11/2020]. Disponível em: <https://www.gov.br/anvisa/pt-br/assuntos/alimentos/rotulagem>.
- CHAUDHURI, Arindam *et al.* Optical Character Recognition Systems. *In: OPTICAL Character Recognition Systems for Different Languages with Soft Computing*. Cham: Springer International Publishing, 2017. P. 9–41. ISBN 978-3-319-50252-6. DOI: 10.1007/978-3-319-50252-6_2. Disponível em: https://doi.org/10.1007/978-3-319-50252-6_2.
- DEVLIN, Jacob *et al.* Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.
- GONZALEZ, Alvaro *et al.* Text location in complex images. *In: PROCEEDINGS of the 21st International Conference on Pattern Recognition (ICPR2012)*. [S.l.: s.n.], jan. 2012. P. 617–620.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.
- GUNDIMEDA, Venugopal *et al.* An Automated Computer Vision System for Extraction of Retail Food Product Metadata: AICC 2018. *In: [S.l.: s.n.]*, jan. 2019. P. 199–216. ISBN 978-981-13-1579-4. DOI: 10.1007/978-981-13-1580-0_20.
- JAIN, R. *et al.* **Machine Vision**. [S.l.]: McGraw-Hill, 1995. (Computer Science Series). ISBN 9780070320185.
- KHASGIWALA, Anuj. **Word recognition in nutrition labels with convolutional neural network**. 2018. Tese (Doutorado) – Utah State University.
- KRUSKAL, Joseph B. An overview of sequence comparison: Time warps, string edits, and macromolecules. **SIAM review**, SIAM, v. 25, n. 2, p. 201–237, 1983.
- MATAS, Jiri *et al.* Robust wide-baseline stereo from maximally stable extremal regions. **Image and vision computing**, Elsevier, v. 22, n. 10, p. 761–767, 2004.
- MATSUNAGA, Nate; SULLIVAN, Rick. Image processing for the extraction of nutritional information from food labels. Santa Clara: Santa Clara University, 2015.
- NGUYEN, Thi Tuyet Hai *et al.* Neural Machine Translation with BERT for Post-OCR Error Detection and Correction. *In: PROCEEDINGS of the ACM/IEEE Joint Conference on Digital Libraries in 2020. Virtual Event, China: Association for Computing Machinery*, 2020. (JCDL '20), p. 333–336. DOI: 10.1145/3383583.3398605. Disponível em: <https://doi.org/10.1145/3383583.3398605>.

PATEL, Chirag; PATEL, Atul; PATEL, Dharmendra. Optical Character Recognition by Open source OCR Tool Tesseract: A Case Study. **International Journal of Computer Applications**, v. 55, p. 50–56, out. 2012. DOI: 10.5120/8794-2784.

SINGH, Amarjot; BACCHUWAR, Ketan; BHASIN, Akshay. A Survey of OCR Applications. **International Journal of Machine Learning and Computing (IJMLC)**, jan. 2012. DOI: 10.7763/IJMLC.2012.V2.137.

SINGH, Sukhpreet. Optical character recognition techniques: a survey. **Journal of emerging Trends in Computing and information Sciences**, v. 4, n. 6, p. 545–550, 2013.

SPORICI, Dan; CUȘNIR, Elena; BOIANGIU, Costin-Anton. Improving the Accuracy of Tesseract 4.0 OCR Engine Using Convolution-Based Preprocessing. **Symmetry**, v. 12, p. 715, mai. 2020. DOI: 10.3390/sym12050715.

SZELISKI, Richard. **Computer Vision: Algorithms and Applications**. 1st. Berlin, Heidelberg: Springer-Verlag, 2010. ISBN 1848829345.

TESSERACT. **tessdoc - Tesseract documentation**. [*S.l.: s.n.*], 2021. [Online; acessado em 21-abril-2021]. Disponível em:
<https://tesseract-ocr.github.io/tessdoc/>.

APÊNDICE A – CÓDIGO FONTE

O código fonte desenvolvido neste projeto, em sua integralidade, está disponível no repositório do autor no GitHub, no seguinte link: <https://github.com/cchenzi/tcc>. Lá está descrito como executar tanto os experimentos como a aplicação. Para fins de completude do documento, este apêndice apresentará o código fonte da aplicação desenvolvida, paNNe.

app.py

```

1  import streamlit as st
2  from config import PROJECT_TILE, PROJECT_SUBTITLE, logo_html
3
4  import pandas as pd
5  import pytesseract
6  import glob
7  import cv2
8  from utils import (
9      encode_and_prepare_image,
10     process_text,
11     get_keywords_digits,
12     merge_dicts,
13 )
14 from image_processing import prepare_composition, compose_functions, resize_img
15 from texts import (
16     PROCESS_DEFINITIONS,
17     IMAGE_SELECTION,
18     IMAGE_LOAD,
19     APPLY_ORC,
20     RESIZE_OPTIONS,
21     BLUR_1,
22     THRESHOLDING_TEXT,
23     HISTOGRAM_TEXT,
24     BLUR_2,
25     TEXT_PROCESSING_TEXT,
26     WORDS_TEXT,
27     words_water,
28     words_food,
29 )
30
31
32 languages = {"engUS": "English", "ptBR": "Português"}
33 words_dict = {"Water": words_water, "Food": words_food}
34
35
36 st.markdown(
37     logo_html,
38     unsafe_allow_html=True,

```

```
39 )
40 st.title(PROJECT_TILE)
41
42 lang = st.sidebar.selectbox(
43     "",
44     ["engUS", "ptBR"],
45     format_func=lambda x: languages.get(x),
46 )
47 st.header(PROJECT_SUBTITLE[lang])
48 st.sidebar.title(PROCESS_DEFINITIONS[lang])
49 filenames = list(sorted(glob.glob("./images/sample_sets/4/*")))
50 file_to_show = st.sidebar.selectbox(IMAGE_SELECTION[lang], filenames)
51
52 load_image = st.sidebar.checkbox(IMAGE_LOAD[lang], value=False)
53 st.sidebar.subheader(APPLY_ORC[lang])
54 ch_o = st.sidebar.checkbox("Original", value=False)
55
56
57 st.sidebar.subheader("Process image with...")
58 resize_options = st.sidebar.multiselect(
59     RESIZE_OPTIONS[lang],
60     options=[0.1, 0.3, 0.5, 0.7, 0.9, 1],
61     default=[0.1, 0.3, 0.5, 0.7, 0.9, 1],
62 )
63 blurry = st.sidebar.checkbox(BLUR_1[lang], value=False)
64 blurry_value = None
65 bilateral_filtering_sigma = None
66 if blurry:
67     blurry_value = st.sidebar.slider(
68         "Kernel", min_value=1, value=3, max_value=21, step=2
69     )
70 thresholding = st.sidebar.checkbox(THRESHOLDING_TEXT[lang], value=False)
71 contrast = st.sidebar.checkbox(HISTOGRAM_TEXT[lang], value=False)
72 filtering = st.sidebar.checkbox(BLUR_2[lang], value=False)
73 if filtering:
74     bilateral_filtering_sigma = st.sidebar.slider(
75         "Sigma", min_value=5, value=50, max_value=100, step=5
76     )
77
78 functions_to_compose = prepare_composition(
79     blurry, blurry_value, thresholding, contrast, filtering, bilateral_filtering_sigma
80 )
81
82 st.sidebar.write("-----")
83 st.sidebar.subheader(TEXT_PROCESSING_TEXT[lang])
84 words_database = st.sidebar.selectbox("Select dataset", ["Food", "Water"])
85 words_to_consider = st.sidebar.multiselect(
86     WORDS_TEXT[lang],
87     options=words_dict[words_database],
```

```
88     default=words_dict[words_database],
89 )
90
91
92 if not load_image:
93     st.write("Waiting for image...")
94 else:
95     img = cv2.imread(file_to_show, cv2.IMREAD_GRAYSCALE)
96     st.markdown(
97         encode_and_prepare_image("Original", img),
98         unsafe_allow_html=True,
99     )
100    result_dict = {}
101    result_cols = [
102        "scale_factor",
103        "recovered_text",
104        "terms",
105        "terms_to_search",
106        "acc_terms",
107        "acc_tts",
108    ]
109    result_df = pd.DataFrame(columns=result_cols)
110    if ch_o:
111        for scale_factor in resize_options:
112            st.text("-----")
113            resized = resize_img(img, scale_factor)
114            composed_image = compose_functions(resized, functions_to_compose)
115            st.markdown(
116                encode_and_prepare_image(str(scale_factor), composed_image),
117                unsafe_allow_html=True,
118            )
119
120            txt = pytesseract.image_to_string(composed_image, lang="por")
121            try:
122                terms, terms_to_search = process_text(txt, words_to_consider)
123                result = get_keywords_digits(terms_to_search, txt)
124                result_dict = {}
125                result_dict = dict(
126                    zip(
127                        result_cols,
128                        [
129                            str(scale_factor),
130                            txt,
131                            terms,
132                            result,
133                            len(terms),
134                            len(result),
135                        ],
136                )
```

```

137         )
138         result_df = result_df.append(result_dict, ignore_index=True)
139     except:
140         print(f"Unable to process image to {scale_factor}!")
141     st.write("-----")
142     st.header("Results:")
143     result_df = result_df.sort_values(by=["acc_terms", "acc_tts"],
144     ↪ ascending=False)
145     terms_set = list(
146         set([item for sublist in result_df["terms"] for item in sublist])
147     )
148     digits_set = merge_dicts(result_df["terms_to_search"])
149     st.write(result_df)
150     best_result = result_df.iloc[0]
151     st.subheader("Accuracy")
152     st.write(f"Combined result acc = {len(terms_set)/len(words_to_consider)}")
153     st.write(
154         f'Best result for scale_factor = {best_result["scale_factor"]}, with acc =
155         ↪ {best_result["acc_terms"]/len(words_to_consider)}!'
156     )
157     st.subheader("Terms recovered (combined)")
158     st.write(terms_set)
159     st.subheader("Terms recovered (best result)")
160     st.write(best_result["terms"])
161     st.subheader("Digits recovered (combined)")
162     st.write(digits_set)
163     st.subheader("Digits recovered (best result)")
164     st.write(best_result["terms_to_search"])
165     st.write("-----")
166     st.subheader("Recovered text...")
167     st.write(best_result["recovered_text"])

```

config.py

```

1 PROJECT_TILE = "paNNe"
2 PROJECT_SUBTITLE = {
3     "ptBR": "Leitor de Rótulos de Embalagens",
4     "engUS": "Package Label Scanner",
5 }
6
7 logo = '</img>'
10 logo_html = '<div style="text-align: right;">' + logo + '</div>'

```

image_processing.py

```
1 import cv2
2
3
4 def prepare_composition(
5     blurry, blurry_value, thresholding, contrast, filtering, bilateral_filtering_sigma
6 ):
7     funcs = []
8     if blurry:
9         funcs.append((remove_noise, {"kernel": blurry_value}))
10    if thresholding:
11        funcs.append((apply_otsu_thresholding, {}))
12    if contrast:
13        funcs.append((histogram_equalization, {}))
14    if filtering:
15        funcs.append((bilateral_filtering, {"sigma": bilateral_filtering_sigma}))
16    return funcs
17
18
19 def compose_functions(image, funcs):
20     for (func, args) in funcs:
21         image = func(image, **args)
22     return image
23
24
25 def remove_noise(image, kernel):
26     # print("Doing Gaussian Blur")
27     return cv2.GaussianBlur(image, (kernel, kernel), 0)
28
29
30 def apply_otsu_thresholding(image):
31     # print("Doing Thresholding with Otsu")
32     return cv2.threshold(image, 0, 255, cv2.THRESH_OTSU)[1]
33
34
35 def histogram_equalization(image):
36     # print("Doing Histogram Equalization")
37     return cv2.equalizeHist(image)
38
39
40 def bilateral_filtering(image, sigma):
41     # print("Doing Bilateral Filtering")
42     return cv2.bilateralFilter(image, 9, sigma, sigma)
43
44
45 def resize_img(img, scale_percent):
46     width = int(img.shape[1] * scale_percent)
47     height = int(img.shape[0] * scale_percent)
48     dim = (width, height)
```

```
49     return cv2.resize(img, dim, interpolation=cv2.INTER_AREA)
50
51
52 def e4_all_funcs(image):
53     funcs = prepare_composition(True, 3, True, True, True, 50)
54     return compose_functions(image, funcs)
```

texts.py

```
1 PROCESS_DEFINITIONS = {
2     "ptBR": "Definição de processamento",
3     "engUS": "Processing definitions",
4 }
5
6 IMAGE_SELECTION = {"ptBR": "Selecione uma imagem", "engUS": "Select a image"}
7
8 IMAGE_LOAD = {"ptBR": "Carregar imagem", "engUS": "Load image"}
9
10 APPLY_ORC = {"ptBR": "Aplicar OCR...", "engUS": "Apply OCR..."}
11
12 RESIZE_OPTIONS = {
13     "ptBR": "Fatores de escala para redimensionamento",
14     "engUS": "Scale factors to resizing",
15 }
16
17 BLUR_1 = {"ptBR": "Suavização Gaussiana", "engUS": "Gaussian Blur"}
18 THRESHOLDING_TEXT = {
19     "ptBR": "Limiarização (Otsu)",
20     "engUS": "Thresholding (Otsu)",
21 }
22
23 HISTOGRAM_TEXT = {
24     "ptBR": "Contraste (equalização de histograma)",
25     "engUS": "Contrast (histogram equalization)",
26 }
27
28 BLUR_2 = {
29     "ptBR": "Suavização com Filtragem Bilateral",
30     "engUS": "Bilateral Filtering",
31 }
32
33 TEXT_PROCESSING_TEXT = {
34     "ptBR": "Definição do processamento de texto",
35     "engUS": "Text processing definitions",
36 }
37
38 WORDS_TEXT = {
39     "ptBR": "Palavras a serem consideradas",
```



```
40     "engUS": "Words to consider",
41 }
42
43
44 words_water = [
45     "bicarbonato",
46     "sodio",
47     "calcio",
48     "magnesio",
49     "cloreto",
50     "sulfato",
51     "potassio",
52     "nitrato",
53     "fluoreto",
54     "estroncio",
55     "bario",
56 ]
57 words_food = [
58     "carboidratos",
59     "proteinas",
60     "gorduras totais",
61     "gorduras saturadas",
62     "gorduras trans",
63     "fibra alimentar",
64     "sodio",
65     "valor energetico",
66 ]
```

utils.py

```
1 import cv2
2 import jiwer
3 import Levenshtein as lev
4 import pandas as pd
5 import base64
6 import re
7
8
9 def get_keywords_digits(terms, text):
10     result = {}
11     rerere = re.compile(r"\d+(,\d+)*")
12     default_transformation = jiwer.Compose(
13         [
14             jiwer.ToLowerCase(),
15             jiwer.RemoveWhiteSpace(replace_by_space=True),
16             jiwer.RemoveMultipleSpaces(),
17             jiwer.Strip(),
18             jiwer.SentencesToListOfWords(word_delimiter=" "),
```

```
19         jiwer.RemoveEmptyStrings(),
20     ]
21 )
22 transformed_text = default_transformation(text)
23
24 words_to_search = list(terms.keys())
25
26 for word in words_to_search:
27     begin = -1
28     end = -1
29     found = False
30     gram_len = len(word.split(" "))
31     for i in range(0, len(transformed_text) - gram_len + 1):
32         instance = " ".join(transformed_text[i : i + gram_len])
33         instances_to_consider = [
34             " ".join(transformed_text[i : i + aux])
35             for aux in range(1, gram_len + 1)
36         ]
37         if word == instance:
38             begin = i
39             found = True
40         elif (
41             any(
42                 inside_instance in words_to_search
43                 for inside_instance in instances_to_consider
44             )
45             and found
46         ):
47             break
48         end = i + gram_len
49     if begin != -1 and end != -1:
50         line = " ".join(transformed_text[begin:end])
51         possible_number = rerere.search(line)
52         if possible_number:
53             result[terms[word]] = possible_number.group(0)
54     return result
55
56
57 def process_text(text, words):
58     df = pd.DataFrame(columns=["word", *words])
59     default_transformation = jiwer.Compose(
60         [
61             jiwer.ToLowerCase(),
62             jiwer.RemoveWhiteSpace(replace_by_space=True),
63             jiwer.RemoveMultipleSpaces(),
64             jiwer.Strip(),
65             jiwer.SentencesToListOfWords(word_delimiter=" "),
66             jiwer.RemoveEmptyStrings(),
67         ]
```

```
68     )
69     transformed_text = default_transformation(text)
70     dd = {}
71     for word in words:
72         gram_len = len(word.split(" "))
73         for i in range(0, len(transformed_text) - gram_len + 1):
74             instance = " ".join(transformed_text[i : i + gram_len])
75             distance = lev.distance(instance, word)
76             aux_dd = (word, distance)
77             if instance not in dd:
78                 dd[instance] = [aux_dd]
79             else:
80                 dd[instance].append(aux_dd)
81
82     for key, values in dd.items():
83         bla = {}
84         bla["word"] = key
85         for w, d in values:
86             bla[w] = d
87         df = df.append(bla, ignore_index=True)
88
89     df = df.fillna(666)
90     df[words] = df[words].apply(pd.to_numeric)
91     terms_df = df.iloc[df[words].idxmin().values]
92     col_idxxs = terms_df[words].idxmin(axis=1).values
93     terms = []
94     terms_to_search = {}
95     distance_acc = 0
96     for term, probable_word in dict(zip(terms_df.word, col_idxxs)).items():
97         value = terms_df.query("word == @term")[probable_word].values[0]
98         if value <= 2:
99             # print(f"distance between {term} and {probable_word} == {value}")
100             terms.append(probable_word)
101             terms_to_search[term] = probable_word
102             distance_acc += value
103         # else:
104         # print(f"for {term}={probable_word}, not enough={value}")
105     return (list(set(terms)), terms_to_search)
106
107
108 def html_details_session(title, image_str):
109     return f"""
110         <details>
111             <summary>{title}</summary>
112             <img src='data:image/png;base64,{image_str}' width='500' height='600'>.
113         </details>
114     """
115
116
```

```
117 def encode_and_prepare_image(title, image):
118     _, im_png = cv2.imencode(".png", image)
119     byte_img = base64.b64encode(im_png.tobytes()).decode()
120     return html_details_session(title, byte_img)
121
122
123 def merge_dicts(dict_list):
124     result = {}
125     for dd in dict_list:
126         for (k, v) in dd.items():
127             if k not in result:
128                 result[k] = {v}
129             else:
130                 result[k].add(v)
131     for (k, v) in result.items():
132         result[k] = list(v)
133     return result
```

legacy.py

```
1 import pytesseract
2 import cv2
3
4
5 def calculate_rectangle_coordinates(bounding_box):
6     x, y, w, h = bounding_box
7     xmax = x + w
8     ymax = y + h
9     xmin = x
10    ymin = y
11    return ((xmax, ymax), (xmin, ymin))
12
13
14 def calculate_iou(bb_1, bb_2):
15     (xmax_1, ymax_1), (xmin_1, ymin_1) = calculate_rectangle_coordinates(bb_1)
16     (xmax_2, ymax_2), (xmin_2, ymin_2) = calculate_rectangle_coordinates(bb_2)
17
18     x_left = max(xmin_1, xmin_2)
19     y_top = max(ymin_1, ymin_2)
20     x_right = min(xmax_1, xmax_2)
21     y_bottom = min(ymax_1, ymax_2)
22
23     intersection_area = (x_right - x_left) * (y_bottom - y_top)
24     if x_right < x_left or y_bottom < y_top:
25         return 0
26
27     bb1_area = (xmax_1 - xmin_1) * (ymax_1 - ymin_1)
28     bb2_area = (xmax_2 - xmin_2) * (ymax_2 - ymin_2)
```

```
29
30     iou = intersection_area / float(bb1_area + bb2_area - intersection_area)
31
32     return iou
33
34
35 def filter_bounding_boxes(bounding_boxes, threshold=0.9):
36     filtered_bb = [bounding_boxes[0]]
37     for bb in bounding_boxes[1:]:
38         iou = calculate_iou(filtered_bb[-1], bb)
39         if iou >= threshold:
40             filtered_bb.pop()
41
42             filtered_bb.append(bb)
43
44     return filtered_bb
45
46
47 def draw_bounding_boxes(image, bounding_boxes):
48     vis = image.copy()
49     vis = cv2.cvtColor(vis, cv2.COLOR_BGR2RGB)
50     for bb in bounding_boxes:
51         (xmax, ymax), (xmin, ymin) = calculate_rectangle_coordinates(bb)
52         cv2.rectangle(
53             vis, (int(xmin), int(ymin)), (int(xmax), int(ymax)), (255, 255, 0), 2
54         )
55     return vis
56
57
58 def resize_bounding_boxes(bounding_boxes, amount):
59     resized = []
60     for bb in bounding_boxes:
61         (xmax, ymax), (xmin, ymin) = calculate_rectangle_coordinates(bb)
62         xmin = int((1 - amount) * xmin)
63         ymin = int((1 - amount) * ymin)
64         xmax = int((1 + amount) * xmax)
65         ymax = int((1 + amount) * ymax)
66         resized.append([xmin, ymin, xmax - xmin, ymax - ymin])
67     return resized
68
69
70 def merge_boxes(bounding_boxes, threshold=10, amount=0.2, pixels_to_add=30):
71     result = []
72     idx_list = []
73
74     for idx_current, bb_current in enumerate(bounding_boxes):
75
76         if idx_current not in idx_list:
77             (xmax_1, ymax_1), (xmin_1, ymin_1) = calculate_rectangle_coordinates(
```

```

78         bb_current
79     )
80     idx_list.append(idx_current)
81     width = xmax_1 - xmin_1
82     xmax_1 += min(width * amount, pixels_to_add)
83     #             print(f"CURRENT=max({(xmax_1, ymax_1)}), min({(xmin_1,
↪   ymin_1)}))")
84     for idx_candidate, bb_candidate in enumerate(
85         bounding_boxes[idx_current + 1 :])
86     ):
87         (xmax_2, ymax_2), (xmin_2, ymin_2) = calculate_rectangle_coordinates(
88             bb_candidate
89         )
90
91         width = xmax_2 - xmin_2
92         xmin_2 -= min(width * amount, pixels_to_add)
93         xmax_2 += min(width * amount, pixels_to_add)
94
95         x_left = max(xmin_1, xmin_2)
96         y_top = max(ymin_1, ymin_2)
97         x_right = min(xmax_1, xmax_2)
98         y_bottom = min(ymax_1, ymax_2)
99
100        intersection_area = (x_right - x_left) * (y_bottom - y_top)
101        if x_right < x_left or y_bottom < y_top:
102            intersection_area = 0
103
104        #             print(f"CANDIDATE=max({(xmax_2, ymax_2)}),
↪   min({(xmin_2, ymin_2)}))")
105        if intersection_area > threshold:
106            xmax_1 = max(xmax_1, xmax_2)
107            xmin_1 = min(xmin_1, xmin_2)
108            ymin_1 = min(ymin_1, ymin_2)
109            ymax_1 = max(ymax_1, ymax_2)
110            idx_list.append(idx_current + 1 + idx_candidate)
111            #             print(f"UPDATED=max({(xmax_1, ymax_1)}),
↪   min({(xmin_1, ymin_1)}), AREA={intersection_area}")
112        else:
113            break
114
115        result.append([xmin_1, ymin_1, xmax_1 - xmin_1, ymax_1 - ymin_1])
116        #             print(f"LEN={len(result)}, NEW=max({(xmax_1, ymax_1)}), min({(xmin_1,
↪   ymin_1)}), w={xmax_1 - xmin_1}, h={ymax_1 - ymin_1}")
117        #             print("-----")
118
119    return result
120
121
122    def do_all(img):

```

```
123     mser = cv2.MSER_create()
124     regions, bounding_boxes = mser.detectRegions(img)
125     filtered = filter_bounding_boxes(bounding_boxes, threshold=0.7)
126     merged = merge_boxes(filtered, threshold=0, pixels_to_add=100)
127     print(
128         f"ORIGINAL={len(bounding_boxes)}, FILTERED={len(filtered)},
        ↪  MERGED={len(merged)}"
129     )
130     resized = resize_bounding_boxes(merged, 0.01)
131     return filtered, merged, resized
132
133
134 def apply_ocr_e3(image, bounding_boxes):
135     results = []
136     y, x = image.shape
137     for idx, bb in enumerate(bounding_boxes):
138         #         if idx % 50 == 0:
139             #             print(f"NOW IN {idx}... {len(bounding_boxes) - idx} LEFT")
140             (xmax, ymax), (xmin, ymin) = calculate_rectangle_coordinates(bb)
141             ymin = max(0, int(ymin))
142             ymax = min(y, int(ymax))
143             xmin = max(0, int(xmin))
144             xmax = min(x, int(xmax))
145             #             print(x, y, (xmax, ymax), (xmin, ymin))
146             output = pytesseract.image_to_string(
147                 image[int(ymin) : int(ymax), int(xmin) : int(xmax)], lang="por"
148             )
149             results.append(output)
150     return " ".join(results)
```

APÊNDICE B – ARTIGO

Neste apêndice será apresentado o artigo no formato SBC, referente ao presente projeto.

Identificação de propriedades de produtos a partir da leitura automática de rótulos

Francisco Vicenzi¹, Mauro Roisenberg¹, Lucas May Petry¹, Alexandre Gonçalves Silva¹

¹Departamento de Informática e Estatística – Universidade Federal de Santa Catarina (UFSC)
Florianópolis – RS – Brazil

Abstract. *The identification of product properties from the automatic reading of labels can represent a gain in logistics, safety, accessibility, and convenience for all stakeholders. Thus, this project will present an application that can automatically read labels, identify useful attributes, and validate them using predefined rules. To accomplish this objectives, Artificial Intelligence techniques will be used, ranging from the Computer Vision part, through Optical Character Recognition and Machine Learning. Besides the proposed application, a bibliographic review of the State of the Art and metrics to evaluate the performance of the techniques used are also presented.*

Resumo. *A identificação de propriedades de produtos a partir da leitura automática de rótulos pode representar um ganho em logística, segurança, acessibilidade e comodidade para todas as partes envolvidas. Deste modo, este projeto apresentará uma aplicação capaz de ler automaticamente rótulos, identificar atributos que sejam úteis e validá-los utilizando-se de regras pré-definidas. Para isso, serão utilizadas técnicas de Inteligência Artificial, abrangendo desde a parte de Visão Computacional, através do Reconhecimento Óptico de Caracteres e Aprendizado de Máquina. Além da aplicação proposta, também é apresentada uma revisão bibliográfica do Estado da Arte e métricas para avaliar o desempenho das técnicas utilizadas.*

1. Introdução

Rótulos correspondem, em sua maioria, à fonte primária de informação acerca de um produto que um consumidor pode ter. Em relação a produtos alimentícios, ingredientes utilizados, valores nutricionais, prazo de validade são exemplos de informações apresentadas, geralmente, em rótulos. Como a legislação sobre rotulagem varia entre países, é necessário estar atento ao padrão nacional a ser utilizado. No Brasil, o órgão que estabelece e fiscaliza as normas de rotulagem é a Anvisa [Agência Nacional de Vigilância Sanitária - Anvisa 2020]. Deste modo, são definidas, além das informações obrigatórias, maneiras de como dispô-las na impressão do rótulo.

Podemos identificar três agentes principais interessados na rotulagem correta: consumidor, produtor e agência reguladora. É de interesse do consumidor que as informações estejam claramente disponíveis, arranjadas de uma maneira que seja fácil identificar questões importantes para a sua saúde, desde ingredientes que possam ser alergênicos até valores nutricionais específicos para uma dieta. O papel da agência reguladora consiste, em sua essência, estabelecer normas e fiscalizar produtos para que, no mínimo, o consumidor tenha ciência do que esteja consumindo. Já o produtor presta

atenção na rotulagem justamente para seguir as normas previstas, evitando processos judiciais e fomentando uma clientela fidelizada, a partir da qualidade e coerência de seus produtos.

A identificação de propriedades de produtos a partir da leitura automática de rótulos pode representar um ganho para as três partes citadas anteriormente. Aliado a um conjunto de regras, podemos estabelecer e abstrair diversas tarefas a serem realizadas por essa aplicação. Por exemplo, poderia-se carregar as normas da Anvisa e checar, produto por produto, se os rótulos possuem todas as informações necessárias para a comercialização, se estão danificados, etc. Ainda, um consumidor poderia, com seu *smartphone*, tirar uma foto de um produto e verificar se os valores nutricionais correspondem para sua dieta, se possui algum ingrediente que deva prestar atenção, etc. Em relação à acessibilidade, seria possível aliar à aplicação um processo de transcrição de texto para fala, permitindo que pessoas com baixa visão possam ter um maior entendimento do produto em suas mãos.

Neste projeto, utilizaremos técnicas de Visão Computacional (CV) e Aprendizado de Máquina (ML) para desenvolver uma aplicação capaz de identificar propriedades de imagens de rótulos. A identificação automática de propriedade de textos corresponde a um problema complexo [Gonzalez et al. 2012] e recorrente do campo da Visão Computacional. Fatores como diferentes tipos de fontes, cores, tamanhos, ângulos e distorções de imagem, por exemplo, tornam o reconhecimento de texto através de imagens uma tarefa difícil. Neste interím, para a primeira parte da implementação, utilizaremos técnicas de Reconhecimento Ótico de Caracteres (OCR).

O OCR corresponde ao processo de identificar padrões em imagens digitalizadas, fazendo assim a sua correspondência para caracteres pré-definidos (alfanuméricos ou não) [Chaudhuri et al. 2017]. Existem diversas maneiras de se realizar OCR. Entretanto, alguns passos costumam ser comuns aos diversos tipos de projeto: varredura, ou *scan*, da imagem, à procura dos textos; segmentação dos caracteres; extração de *features*; treino dos modelos e classificação dos caracteres; pós-processamento [Singh 2013] [Chaudhuri et al. 2017]. Para etapa de treinamento e classificação, será utilizada uma *engine* de código aberto de OCR, chamada Tesseract. A etapa de pós-processamento consistirá na validação das informações obtidas para com as regras previamente estabelecidas para o sistema, sejam elas legislações, restrições alimentares, etc. Nesta etapa, ainda é possível a utilização de Modelos de Linguagem, técnica esta de Processamento de Linguagem Natural, para melhorar a performance da nossa aplicação. Modelos de Linguagem, como o BERT [Devlin et al. 2018], podem ser ferramentas importantes para minimizar erros oriundos do OCR [Nguyen et al. 2020] ou, até mesmo, apontar erros do próprio rótulo. Estratégias menos robustas, tal como similaridade de texto, também podem ser utilizadas para o pós-processamento.

Portanto, a leitura de rótulos de produtos, ainda que trivial para humanos, consiste em um desafio tecnológico a ser superado. Sua automatização pode trazer diversos benefícios, seja de logística, segurança, acessibilidade ou comodidade. Ao fim deste projeto, apresentaremos uma aplicação que realize a leitura automática de rótulos, identificando atributos que sejam úteis para classificação e validação através de regras pré-definidas.

2. Conceitos Básicos e Trabalhos Relacionados

2.1. Conceitos Básicos

2.1.1. Visão Computacional

A visão computacional representa a área de estudo na qual o principal interesse consiste em desenvolver sistemas e técnicas capazes de obter informações úteis de imagens [Jain et al. 1995]. O ato de observar uma cena e entender o que se passa nela é uma tarefa simples, em geral, para humanos. Desde crianças, é natural que desenvolvamos a capacidade de reconhecer padrões e identificar elementos que façam parte da composição do ambiente. Entretanto, traduzir essa capacidade para computadores é uma tarefa extremamente difícil e que representa décadas de estudo até então.

A figura 1 apresenta uma foto tirada pelo autor em 2019. Pela ótica de um ser humano, facilmente conseguimos descrever a cena: mais à frente, um cachorro dormindo em uma almofada com listras brancas e vermelhas; ao fundo, desfocado, um par de tênis sobre um tapete azul e branco. Já pela ótica de um computador, porém, o que é possível recuperar de informação?

Enquanto a visão humana acontece em três dimensões, as imagens são representadas digitalmente em apenas duas dimensões. Em suma, o objeto de trabalho da visão computacional corresponde a matrizes de pixels que representam imagens. Por exemplo, a figura 2 apresenta a forma matricial resumida da imagem 1, onde cada elemento corresponde aos valores RGB de cada pixel. Também é possível representar imagens de formas mais simples, em que as mais comuns são em escala de cinza e de forma binária. Na escala de cinza, utiliza-se apenas um valor por pixel, que varia de 0 a 255; em imagens binárias, os pixels podem assumir apenas dois valores, 0 e 1.

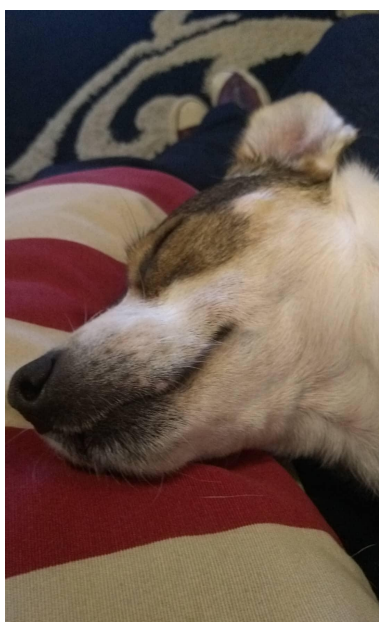


Figure 1. Foto de um cachorro dormindo

$$\begin{bmatrix} [32, 42, 67] & [30, 40, 65] & \cdots & [48, 46, 34] \\ [31, 41, 66] & [29, 39, 64] & \cdots & [43, 41, 29] \\ \vdots & \vdots & \vdots & \vdots \\ [131, 119, 97] & [133, 121, 99] & \cdots & [123, 111, 89] \\ [132, 120, 98] & [133, 121, 99] & \cdots & [122, 110, 88] \end{bmatrix}$$

Figure 2. Representação matricial resumida da imagem à esquerda

Portanto, os métodos de visão computacional, ao agir sobre as representações digitais de imagens, almejam, de certa forma, reconstruir as propriedades do mundo real, a fim de descrevê-lo [Szeliski 2010]. Apesar de todas as dificuldades, diversas técnicas já são utilizadas no cotidiano, como, por exemplo, na leitura automática de placas de automóveis; na telemedicina, reconhecimento de objetos, direção autônoma de automóveis, etc.

2.1.2. Reconhecimento Ótico de Caracteres

O Reconhecimento Ótico de Caracteres é uma tarefa de visão computacional que consiste na identificação e recuperação de texto em imagens digitais. Portanto, sistemas de OCR (acrônimo de *Optical Character Recognition*) realizam a conversão de caracteres dispostos em imagens para textos digitais. Essas imagens, por exemplo, podem conter caracteres escritos à mão, inscritos ou impressos [Singh et al. 2012].

Tesseract

O Tesseract é um sistema de Reconhecimento Ótico de Caracteres de código aberto ¹, distribuído a partir da licença Apache 2.0 e financiado pela Google. Atualmente encontra-se na versão 4.1.1 [Tesseract 2021]. Está disponível para várias plataformas e funciona a partir de linha de comando. Além disso, contém APIs definidas para utilização de aplicativos com sistema gráfico ou bibliotecas.

Até a versão 4.0, a arquitetura consistia, principalmente, na análise de componentes conexos, localização de linhas e reconhecimento de palavras [Patel et al. 2012]. A partir da versão 4.0, foi adicionado à arquitetura um sistema baseado em redes neurais LSTM, além de modelos treinados para mais de 120 línguas. Essa mudança apresentou ganhos fortes em acurácia para o sistema de OCR. Desse modo, é possível utilizar os modelos treinados, aprimorá-los ou, até mesmo, treinar modelos próprios.

2.1.3. Métricas

Duas métricas principais são utilizadas para a avaliação quantitativa dos experimentos: distância de Levenshtein (seção 2.1.3) e acurácia (seção 2.1.3).

O reconhecimento ótico de caracteres muitas vezes retorna palavras com erros de ortografia. Em função disso, faz-se necessário definir uma maneira de conferir se uma palavra foi recuperada ou não, mesmo que não seja uma combinação perfeita. Desse modo, a distância de Levenshtein é utilizada para a obtenção de um valor numérico de similaridade entre termos. Com isso, é possível definir um limiar de aceitação para possíveis erros de ortografia oriundos do OCR.

A acurácia é uma métrica que fornece um significado numérico para a realização do experimento. Desse modo, podemos utilizá-la, neste contexto, para verificar o desempenho das técnicas aplicadas.

Distância de Levenshtein

¹Repositório principal do projeto: <https://github.com/tesseract-ocr/tesseract>

A distância de Levenshtein consiste em uma métrica de similaridade entre sequências de caracteres. A distância é computada como sendo o menor número de substituições necessárias para transformar uma sequência A em B [Kruskal 1983]. O algoritmo é definido a partir de três operações fundamentais: substituição, inserção e remoção.

A tabela 1 mostra exemplos de distâncias computadas entre os termos "batata", "batuta" e "banana". A diagonal principal mostra que a distância entre palavras iguais é nula. A distância entre "batata" e "batuta" é igual a 1; uma substituição é necessária para transformar a primeira na segunda. A maior distância encontrada no exemplo dá-se entre os termos "banana" e "batuta". Para esse caso, são necessárias três substituições.

	batata	batuta	banana
batata	0	1	2
batuta	1	0	3
banana	2	3	0

Table 1. Exemplo de distâncias computadas entre palavras

Acurácia

Definição 2.1 Acurácia. Dado uma lista de termos recuperados R e uma lista de palavras-chave K , a acurácia é computada como

$$Acc(R, K) = \frac{\#R}{\#K}.$$

2.2. Trabalhos Relacionados

A identificação automática de texto em imagens é um tema bastante estudado na literatura. Como a leitura ótica de caracteres já é discutido há décadas, uma das maneiras de se buscar melhores resultados é trabalhar no pré-processamento da imagem para, posteriormente, aplicar métodos já existentes de OCR.

Em geral, os algoritmos de reconhecimento de texto, quando utilizados de forma generalizada, podem ter problemas ao lidar com elementos distintos nas imagens, que não sejam caracteres. Assim, [Gonzalez et al. 2012] apresenta um método de auxílio ao reconhecimento ótico de caracteres. O método proposto abrange a etapa anterior ao reconhecimento: a localização de texto. A segmentação do texto do restante da imagem faz com que a qualidade do processamento e do reconhecimento do texto seja superior à apenas executar os métodos de OCR indiscriminadamente. A metodologia proposta é composta por três etapas: segmentação da imagem para possíveis caracteres; análise baseada em componentes conectados para geração de *features*; e, por fim, classificação de linhas de texto. Na primeira etapa, o algoritmo MSER é utilizado para a identificação de regiões prováveis de se conter textos. São geradas 8 features na segunda etapa do método, como, por exemplo, a proporção e a taxa de ocupação da imagem. A terceira etapa consiste no agrupamento dos caracteres em linhas, utilizando-se de algoritmos de limiarização adaptativos.

Enquanto técnicas mais gerais são desenvolvidas, a popularidade e qualidade do motor de OCR Tesseract fez com que métodos fossem desenvolvidos direcionados a ele.

O método proposto por [Sporici et al. 2020] tem como objetivo melhorar o desempenho do Tesseract a partir do pré-processamento da imagem. Uma das principais contribuições do método é mostrar uma forma de contornar as limitações conhecidas do motor, tais como lidar com possíveis imperfeições da imagem, que fazem com que a qualidade do resultado decaia muito. Neste método, o pré-processamento da imagem é realizado a partir da geração de kernels de convolução, de maneira não supervisionada. Desse modo, são realizados ciclos de treino, onde a imagem com o kernel aplicado é levada ao Tesseract e obtido uma pontuação. Os resultados mostraram-se bastante promissores, sendo que os autores apresentaram métricas como *Character Error Rate (CER)*, *Word Error Rate (WER)*, *Character Accuracy (CAR)*, *Precision*, *Recall* e *F1 Score*.

O método proposto por [Khasgiwala 2018] realiza a identificação de rótulos nutricionais em embalagens de alimentos utilizando Redes Neurais Convolucionais. Neste trabalho, cada linha do rótulo é segmentada utilizando o algoritmo TSAW (*Text Skew Angle Wavelets*). Desse modo, busca-se extrair cada elemento do rótulo e seu valor associado. Posteriormente, cada linha é levada à Rede Neural Convolutiva para identificar o elemento segmentado. Além disso, o Tesseract também é utilizado para realizar o OCR de cada segmento, a fim de comparar os resultados. Um dos pontos importantes desse trabalho consiste na segmentação das linhas do rótulo, assim como a possibilidade de receber imagens que sejam rotacionadas. Os resultados reportam acurácia máxima de aproximadamente 69%. Nesse sentido, são descritas como razões para a baixa acurácia o tamanho do conjunto de dados, ruído no conjunto de dados, dimensão das imagens, entre outras.

[Matsunaga and Sullivan 2015] apresenta uma arquitetura de trabalho baseada em três módulos para a extração de informação nutricional de rótulos de alimentos: pré-processamento da imagem, OCR e pós-processamento do texto. No pré-processamento da imagem, são utilizadas técnicas de visão computacional clássica a fim de extrair o rótulo do restante da imagem. Para isso, utiliza-se suavização, aumento do contraste, filtragem e limiarização para preparar a extração. Essas etapas permitem uma melhor identificação do contorno do rótulo, útil para a extração. O rótulo extraído é aplicado ao Tesseract; sua saída é, posteriormente, tratada com algumas técnicas de correção de ortografia. No final, são obtidos pares de termos e seu valor nutricional. A acurácia deste método consiste na porcentagem de termos nutricionais corretamente identificados no conjunto de dados. O método foi testado em conjuntos de dados com características diferentes, como, por exemplo, rotação, cor e iluminação. A maior parte dos testes teve uma acurácia de aproximadamente 80%.

[Gundimeda et al. 2019] propõe um sistema de extração automática de elementos de rótulo de alimentos baseado em três subsistemas: remoção de fundo da imagem, classificação automática da qualidade da imagem e extração de atributos. O subsistema de remoção de fundo da imagem é composto por técnicas de processamento de imagem para a correção de iluminação, limiarização e projeção. A classificação automática da qualidade da imagem é composta pela classificação entre duas métricas desse subsistema. A primeira é obtida através da detecção de texto de uma área de interesse utilizando análise de regiões e classificador baseado em regras; a segunda é obtida através do OCR da área de interesse. Posteriormente, essas métricas são analisadas para identificar se a imagem é aceitável, rejeitável ou necessita intervenção manual. O subsistema

de extração de atributos busca recuperar elementos como o nome da marca, o título do produto, informações nutricionais, entre outras. Todas as etapas de extração passam pela segmentação da imagem e a aplicação de OCR. A identificação de palavras-chave determina qual fluxo de execução para extração de atributo será utilizado, além de definir a necessidade de pós-processamento do texto obtido. Os resultados consolidados do sistema proposto chegaram, para cada extração de atributo, em uma acurácia de mais de 95%.

3. Desenvolvimento

3.1. Conjunto de Dados

Para a realização dos experimentos, foi confeccionado um conjunto de dados de imagens de garrafas de água mineral e de produtos alimentícios. Em busca de uma maior diversidade, foram selecionadas quatro marcas diferentes de água mineral e outros três produtos alimentícios. Ainda assim, as imagens foram separadas em quatro grupos para os experimentos, dadas características como resolução e ângulo das fotos. Esses grupos foram nomeados em ordem alfabética, de A a D, com intuito de diferenciar sua nomenclatura dos experimentos.

O grupo A do conjunto de dados é composto por imagens que capturam o rótulo junto ao formato original das garrafas. Fazem parte deste primeiro grupo imagens com o corpo inteiro da garrafa, assim como também alguns recortes mais focados no rótulo. A figura 3 mostra três exemplos de imagens para o primeiro caso, enquanto a figura 4 apresenta um exemplo para o segundo.



Figure 3. Rótulos com ângulo original

Para o grupo B do conjunto de dados, os rótulos foram extraídos das garrafas e colocados sob uma superfície plana. Desse modo, a relação de ângulo é modificada, uma vez que perde-se a especificidade do formato do objeto. Na figura 5 é possível observar um exemplo deste grupo.

Para o grupo C do conjunto de dados, foram extraídos componentes menores a partir das figuras do segundo conjunto de dados. A figura 6 apresenta dois componentes recortados da figura 5.

O grupo D de conjunto de dados é composto por rótulos de três produtos alimentícios diferentes e de algumas imagens de água do terceiro conjunto de dados. Os produtos selecionados foram leite, creme de leite e erva-mate. A partir das fotos, os



Figure 4. Rótulo com ângulo original, aproximado



Figure 5. Rótulo planificado, sobre superfície plana

rótulos foram recortados manualmente de forma digital. A figura 7 apresenta duas imagens presentes nesse conjunto.

O sumário do conjunto de dados é apresentado na tabela 2, dividido pelos grupos, junto às suas características previamente discutidas e suas respectivas quantidades.

Classificação: Água Mineral Fluoretada e Fracamente Radioativa na Fonte. Análise Química Boletim nº 212/LAMIN/18 de 25/07/18 LAMIN/CPRM.
NÃO CONTEM GLUTÊN

Composição química (mg/L)	
BICARBONATO	23,13
SÓDIO	2,707
CÁLCIO	2,849
MAGNÉSIO	2,350
CLORETO	2,150
SULFATO	0,690
POTÁSSIO	0,617
NITRATO	0,820
FLUORETO	0,040
ESTRÔNCIO	0,017
BÁRIO	0,014

7 898228 200073

Características Físico-Químicas	
pH a 25 °C	6,60
Temperatura da Água na Fonte	16,7°C
Condutividade Elétrica a 25° C	43,9µS/cm
Resíduo de Evaporação a 180 °C, calculado	40,35mg/L
Radiatividade na Fonte a 20°C e 760mmHg	5,08 Maches

Conservar ao abrigo do sol, em local limpo, seco, arejado e sem odor. **CONCESSIONÁRIA:** ESTÂNCIA HIDROMINERAL SANTA RITA DE CÁSSIA LTDA. CNPJ 03489027/0001-88. BR 282 Km 53 - Caixa Postal 15 - Rio dos Bugres - Rancho Queimado - SC - CEP 88470-000. Processo DNPM nº 815.285/99. Portaria de Lava nº 502. Publicado no DOU 10/10/02. Indústria Brasileira.

FONTA ÁGUA AZUL
Local da fonte: Águas Mornas - SC
www.aguasantarita.com.br
SAC: (48) 3275-0353 | sac@guasantarita.com.br

Embalagem Descartável
Validade: 6 meses
Data de Fabricação:
Vide marcação na embalagem

Figure 6. Componentes extraídos de um rótulo planificado

INFORMAÇÃO NUTRICIONAL
Porção de 15 g (1 colher e ½ de sopa)

Quantidade por porção	% VD (*)	
Valor energético	26 kcal = 109 kJ	1
Carboidratos	0,6 g	0
Proteínas	0	0
Gorduras totais	2,6 g	5
Gorduras saturadas	1,6 g	7
Gorduras trans	0	**
Fibra alimentar	0	0
Sódio	12 mg	1

* % Valores Diários com base em uma dieta de 2.000 kcal ou 8.400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas.
** VD não estabelecido.

INGREDIENTES: leite integral, estabilizantes: trifosfato de sódio (INS 451), citrato de sódio (INS 331iii), monofosfato monossódico (INS 339i) e difosfato dissódico (INS 450i).
ALÉRGICOS: CONTEM LEITE.
CONTÉM LACTOSE. NÃO CONTEM GLUTÊN.

INFORMAÇÃO NUTRICIONAL		
Porção de 200 ml (1 copo)		
Quantidade por porção	% VD (*)	
Valor Energético	116 kcal = 487 kJ	6
Carboidratos	9,1 g	3
Proteínas	6,3 g	8
Gorduras Totais	6,0 g	11
Gorduras Saturadas	4,3 g	20
Gorduras Trans	0 g	**
Fibra Alimentar	0 g	0
Sódio	152 mg	6
Cálcio	209 mg	21

* % Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** Valor não especificado.

MANTENHA EM LOCAL SECO E AREJADO. APÓS ABERTURA, CONSERVAR SOB REFRIGERAÇÃO DE +1 °C A +10 °C E CONSUMIR EM ATÉ 48 HORAS. NÃO É NECESSÁRIO FEVER.

Figure 7. Componentes extraídos de um rótulo plano

Grupo	Características	Quantidade
A	Corpo inteiro da garrafa, ângulo e formato original mantidos, grande resolução	44
B	Rótulos planificados, grande resolução	26
C	Recortes de elementos dos rótulos planificados, baixa resolução	13
D	Recortes de elementos dos rótulos planificados, grande resolução	10

Table 2. Sumário do conjunto de dados

3.2. Experimentos

Os experimentos realizados nesta etapa tiveram caráter exploratório e de análise quantitativa. Foram realizados em um computador com 4 núcleos i7-75000 CPU 2.70GHZ, com 7.6 Gib de memória RAM, placa de vídeo NVIDIA GeForce GT 620M e sistema operacional Manjaro, na versão do kernel 5.9. O código dos experimentos foi implementado em linguagem Python (v. 3.9), com auxílio principalmente das bibliotecas opencv-python (v. 4.5.1), pytesseract (v. 0.3.7) e numpy (v. 1.19.4). Para capturar a ordem cronológica e curva de aprendizado, dividiu-se os experimentos em cinco grupos principais.

3.2.1. Primeiro experimento

Descrição

No primeiro experimento, o objetivo foi a familiarização com o Tesseract, além de buscar entender melhor o problema a ser resolvido. Pode-se dizer que foram realizados testes iniciais exploratórios, uma vez que as imagens foram aplicadas naturalmente, sem nenhum tratamento, no sistema de reconhecimento ótico. Além disso, as imagens utilizadas neste momento foram as do grupo A do conjunto de dados. A ideia principal consistiu em observar qual seria a resposta neste caso, conferir como seria a performance do sistema com as características apresentadas pelas imagens. A figura 8 apresenta um fluxograma que sumariza o primeiro método experimentado.

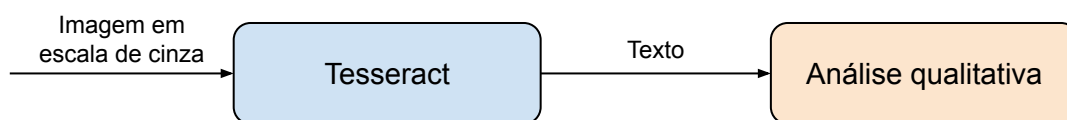


Figure 8. Fluxograma do primeiro método, primeiro e terceiro experimentos

Resultados

Os primeiros resultados não foram satisfatórios, mas serviram como base para pensarmos os próximos experimentos. Em diversas imagens, não foi possível recuperar nenhum texto, apenas algumas letras (ou combinações) sem muito sentido. Entretanto, algumas imagens foram capazes de produzir alguns resultados iniciais, ainda que um número reduzido delas. Em termos de texto, o resultado obtido é apresentado na tabela 3.

Imagem	Resultado
Original, em escala de cinza	WOCONTEM (COMPOSIGAO (CARACTERGTERY GLUTEN. —QUIMICA (mg/L) FISICO-UIMICAs Estrineo..... 0,037 CAICIO s, 25 : Magnésio..... 0,09 g POtS8I...vc... 0,21 [S0010....core. 75,81 ” Vanédo.....,.....mW Su!fato.....fi,28’ e 81,10

Table 3. Resultado obtido no primeiro grupo de experimentos.

3.2.2. Segundo experimento

Descrição

O segundo grupo de experimentos também foi realizado com o primeiro conjunto (A) de dados. Diferente do primeiro experimento, neste momento tentou-se realizar algumas etapas de pré-processamento nas imagens antes de aplicá-las ao Tesseract. As técnicas utilizadas foram de abertura morfológica, remoção de ruído utilizando filtro de mediana, limiarização e Canny.

Uma série de experimentos foi realizada, combinando as técnicas e experimentando alguns parâmetros. As etapas mencionadas são apresentadas na figura 9.



Figure 9. Fluxograma do segundo método, segundo experimento

Resultados

A experimentação em relação à variação dos parâmetros utilizados pelas técnicas mencionadas não foi muito extensa. Após alguns testes, a configuração deste segundo experimento foi a seguinte: para aplicação da técnica de abertura morfológica, utilizou-se um kernel de 5×5 , composta apenas por números 1; para a remoção de ruído, foi utilizado um filtro mediano de 5×5 ; já para a limiarização, utilizou-se o método de Otsu, que adapta o limiar conforme a imagem; para Canny, os valores de limiar foram 10 e 200.

No geral, como os parâmetros não foram escolhidos de forma tão criteriosa, não obteve-se resultados concretos, em geral, neste experimento. Os resultados para esse estão apresentados na tabela 4. Nota-se que não houve melhora no resultado com as técnicas obtidas; de fato, foram até piores do que com a imagem original, apenas em escala de cinza.

Imagem	Resultado
Original, em escala de cinza	WOCONTEM (COMPOSIGAO (CARACTERGTERY GLUTEN. —QUIMICA (mg/L) FISICO-UIMICAs Estrineo..... 0,037 CAICIO s, 25 : Magnésio..... 0,09 g POtS8L...vc... 0,21 [S0010....core. 75,81 ” Vanédo.....mW Su!fato.....fi,28’ e 81,10
Aplicado remoção de ruído	.u’ - MOCONTEM COMPOSICAO (CARACTERGTE: GLUTEN. QUIMICA (mgL) (FisICo. : Estrinc..... 0,037 ’ \\ CACIB it 1,25 \\ Magnésio..... 0,09 \\ = Potéssio..... 0,21 \\ ’— 30di..... 15,81 \\ e TIT \\ ’ OS .61, \\ :3’ - — Bicarbonato....
Aplicado abertura, remoção de ruído e Canny	iip CONTIER fi@@?fi@@@fi?fi@f BLUTER QUIEMICA (mgL) %ffi@&%@%m@ - V= L il O fl @@afi&@z@@ @E?%%@@@WW@%@@% BT i CEeL g, — YO0 e e A 7 [l e B R (010} S o — ol CGE SI N N T B @ F=a IR TWTRE el ANIE N Ly S e fii@ AR R U SNTRe . s 7 A e < // % <e

Table 4. Resultado de uma imagem do segundo experimento, dividido pelas técnicas aplicadas

3.2.3. Terceiro experimento

Descrição

O terceiro grupo de experimentos foi realizado com o terceiro grupo (C) do conjunto de dados. Em relação à metodologia aplicada, este experimento utilizou o método apresentado no primeiro experimento (fluxograma 8), uma vez que as imagens foram levadas ao sistema de OCR sem tratamento algum, apenas em escala de cinza. Entretanto, pela natureza das imagens deste conjunto, foi possível obter resultados satisfatórios.

Resultados

A tabela 5 apresenta três exemplos de resultados obtidos. Para a primeira imagem, observa-se que foi possível obter o texto quase que integralmente, salvo algumas exceções de acentuação e caracteres especiais. A segunda imagem também apresentou um bom resultado; em geral, a maior parte das informações foi recuperada, com exceção da última linha. Já para a terceira imagem, verifica-se que a informação recuperada não foi tão boa. É possível notar que apenas a parte esquerda do rótulo foi previamente coberta, enquanto pouca coisa útil foi recuperada da direita.

3.2.4. Quarto experimento

Descrição

Para o quarto grupo de experimentos, foi implementado uma série de passos de pré-processamento, inspirados na técnica apresentada por [Gonzalez et al. 2012]. Primeiramente, aplicou-se o algoritmo de MSER, a fim de obter regiões de interesse na imagem. Após, foi necessário realizar um filtro sobre as regiões obtidas, uma vez que a implementação utilizada gera um número muito alto, com várias repetições. Para a filtragem, as imagens foram comparadas par a par, sequencialmente, a partir do valor de intersecção sobre a união (IoU). Utilizou-se um limiar de 70% para decidir se as regiões eram iguais. A etapa de filtragem foi realmente necessária; em alguns casos, foi capaz de reduzir aproximadamente 80% as regiões obtidas no primeiro passo. Como o objetivo consiste na identificação de palavras, foi necessário realizar a junção das regiões, a fim de aplicar, posteriormente, o motor de OCR. A junção das regiões deu-se, assim como na filtragem, sequencialmente, par a par. A cada passo, realizava-se uma leve expansão da região e verificava-se a intersecção horizontal entre elas; caso houvesse, eram unidas. Essa etapa de junção também ajudou a reduzir bastante o número de regiões. Por fim, para cada região obtida, realizou-se o seu recorte da imagem original e aplicou-se o sistema de OCR. O fluxograma deste experimento é apresentado na figura 10, enquanto a figura 11 apresenta um exemplo prático destas etapas.

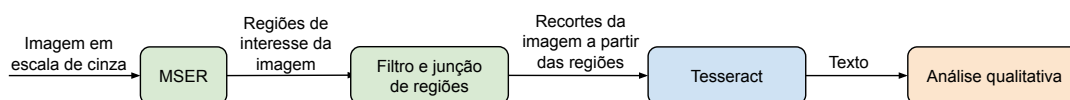


Figure 10. Fluxograma do terceiro método, quarto experimento

Imagem	Resultado Obtido
	<p>CONCESSIONARIA: MINERADORA DE AGUAS RAINHA LTDA (NP):01.310.808/0001-38. INDUSTRIA BRASILEIRA. Distribuido e Comercializado por: Aimaleve Bebidas e Logistica - (NPJ: 18.119.353/0002-51 Avenida Brasil, 804 - Jardim Paran - CEP: 85935-000 Assis Chateaubriand - PR,FONTE BELA VISTA, LOCAL DA FONTE: ESTRADA DA MINA, KM 02, SITIO BELA VISTA, GLEBA VALENCIA, CEP: 86.750-000, IGUARACU - PR. PROCESSO Ne 826.201/1994 DNPM. PORTARIA DE LAVRA Ne 576 DE 18/12/1998, D.O.U DE 21/12/1998., BOLETIM Ne 073/LAMIN/CPRM/2018 DE 18/04/2018.</p>
	<p>Classificagio: Agua Mineral Fluoretada e Fracaimente Radioativa na Fonte. Andlise Quimica Boletim nº 212/LAMIN/18 de 25/07/18 LAMIN/CPRM, NAO CONTEM GLUTEN</p> <p>Composicao quimica (mg/L)</p> <p>BICARBONATO 2313 sÓDIO 2,707 CALCIO 2,849 MAGNESIO 2,350 CLORETO 2,150 SULFATO 0,690 POTASSIO 0,617 NITRATO 0,820 FLUORETO 0,040 0,017 BARIO,w “ II ”J 1 ¥ 9822 z</p>
	<p>NAO CONTEM GLUTEN.</p> <p>CONSERVAR AO ABRIGO DO SOL, EM LOCAL LIMPO, SECO E SEM ODOOR.</p> <p>Temperatura SAC-453125.5000 da 4gua na fonte Emil: sac@grupopereirabrasil com 25,2°C ? i Condutividade Data de envase, validade e . : h elétricaa 25°C Wmfig,“m.mvagd;”dem i 327 pSlem</p> <p>Residuo de evaporagao 1 2180°C, calcul Css 220,17 mg/L</p>

Table 5. Resultados obtidos no terceiro grupo de experimentos

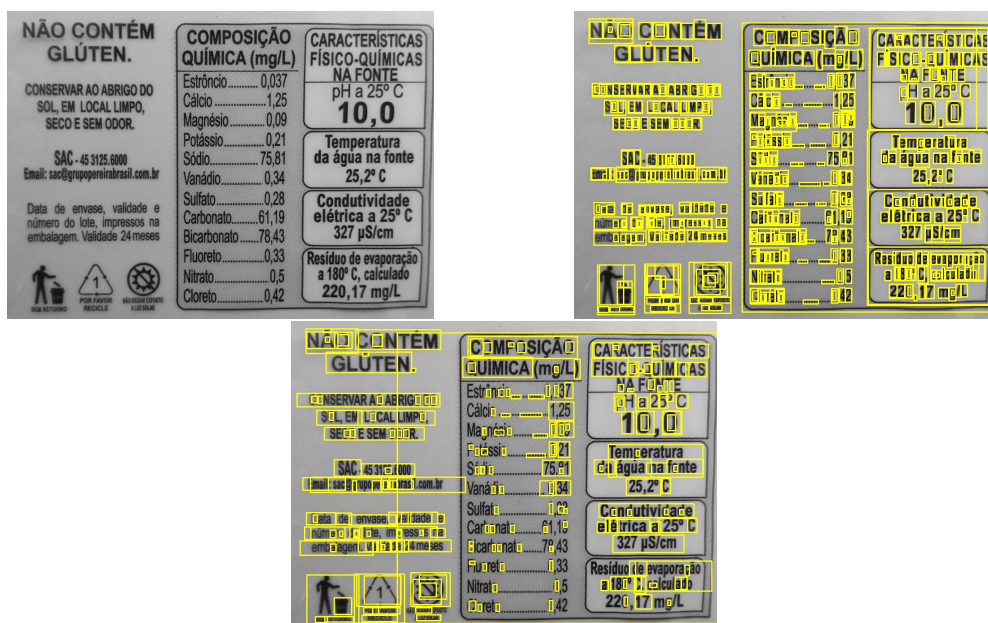


Figure 11. Exemplo de processo do experimento quatro. Na esquerda superior, a imagem original em escala de cinza. Na direita superior, a imagem após MSER e filtragem de regiões. Na parte inferior centralizada, a imagem após junção das regiões.

Resultados

Um exemplo de resultado deste quarto grupo de experimentos é apresentado na tabela 6. Entretanto, observa-se que os resultados não foram tão satisfatórios, tanto na parte qualitativa como na parte de desempenho; aplicar o motor a cada região obtida foi um processo um tanto quanto demorado. Nota-se que algumas regiões necessárias, como na parte de composição química, foram removidas erroneamente. Ainda assim, ao contrário dos experimentos anteriores, foi possível capturar uma maior quantidade de regiões em uma imagem com bastante informação.

3.2.5. Quinto experimento

Descrição

O quinto grupo de experimentos foi realizado com o quarto conjunto (D) de dados. Nesse experimento, além do resgate das etapas de pré-processamento de imagem, foram adicionados passos de redimensionamento e de pós-processamento, permitindo uma análise quantitativa do resultado. O fluxograma 12 sumariza esse experimento.

Primeiramente, a imagem do rótulo é lida em escala de cinza. Após, a imagem redimensionada é levada a uma etapa de pré-processamento com técnicas de visão computacional clássica. Essa etapa de pré-processamento tem como objetivo realçar os caracteres do rótulo ao mesmo tempo que busca remover elementos da imagem que não sejam úteis para o reconhecimento ótico. Neste contexto, quatro técnicas são aplicadas à imagem.

A primeira técnica de tratamento de imagem é a suavização Gaussiana, para

Imagem	Resultado Obtido																								
 <p>NÃO CONTÉM GLÚTEN.</p> <p>CONSERVAR ABRIGADO, SECO, EM LOCAL LIMPO, SECO E SEM DOR.</p> <p>SAC: 45 31 0000</p> <p>email: sac@grupopereira.com.br</p> <p>Data de envase, validade e número lote, impressa na embalagem. Validade 24 meses</p> <p>COMPOSIÇÃO QUÍMICA (mg/L)</p> <table border="1"> <tr><td>Estrôncio</td><td>0,137</td></tr> <tr><td>Cálcio</td><td>1,25</td></tr> <tr><td>Magnésio</td><td>0,09</td></tr> <tr><td>Potássio</td><td>0,21</td></tr> <tr><td>Sódio</td><td>75,81</td></tr> <tr><td>Vanádio</td><td>0,34</td></tr> <tr><td>Sulfato</td><td>1,09</td></tr> <tr><td>Carbonato</td><td>21,19</td></tr> <tr><td>Bicarbonato</td><td>78,43</td></tr> <tr><td>Fluoreto</td><td>0,33</td></tr> <tr><td>Nitrato</td><td>0,5</td></tr> <tr><td>Cloreto</td><td>0,42</td></tr> </table> <p>CARACTERÍSTICAS FÍSICO-QUÍMICAS NA FONTE</p> <p>pH a 25° C: 10,0</p> <p>Temperatura da água na fonte: 25,2° C</p> <p>Condutividade elétrica a 25° C: 327 µS/cm</p> <p>Resíduo de evaporação a 180° C, calculado: 220,17 mg/L</p>	Estrôncio	0,137	Cálcio	1,25	Magnésio	0,09	Potássio	0,21	Sódio	75,81	Vanádio	0,34	Sulfato	1,09	Carbonato	21,19	Bicarbonato	78,43	Fluoreto	0,33	Nitrato	0,5	Cloreto	0,42	<p>— LOCAL LIMPO, pereira.com.br sacOgrupoperei é > — E TEM N., RIGO DO LIMPO, DOR.</p> <p>000 brasil.com.br validade e Dressos nNa 2 24meses &</p> <p>COMPOSIÇÃO (CARACTERÍSTICAS)) QUÍMICA (mg/L) FÍSICO QUÍMICAS POtássio.....m.... 0,21 Temperatura Sin 7581 — da água na fonte Vanádio.....m.... 0,34 A ————— Condutividade elétrica a 25° C 327 uS/cm Resíduo de evapo a 180° C, calcul 220 17 mg/L COMPOSIÇÃO M o E 1,252/ pS/cm UIM Temperatura 3 água na fonte ic água —</p>
Estrôncio	0,137																								
Cálcio	1,25																								
Magnésio	0,09																								
Potássio	0,21																								
Sódio	75,81																								
Vanádio	0,34																								
Sulfato	1,09																								
Carbonato	21,19																								
Bicarbonato	78,43																								
Fluoreto	0,33																								
Nitrato	0,5																								
Cloreto	0,42																								

Table 6. Resultados obtidos no quarto grupo de experimentos

remoção de possíveis ruídos na imagem. Após, é aplicado limiarização adaptativa com o método de Otsu. Em seguida, para aumentar o contraste das regiões da imagem, é aplicado a técnica de equalização de histograma. Por fim, também com intuito de remover possíveis ruídos, é aplicado a técnica de filtragem bilateral. Essa última técnica de suavização, em contraste com a primeira, tem seu principal diferencial no resguardo das bordas, fator importante para o reconhecimento dos caracteres.

Após as etapas de tratamento, a imagem é levada ao sistema de OCR. O texto obtido nesta etapa é aplicado a algumas etapas de processamento a fim de facilitar sua avaliação. Nesse processamento, todas as letras do texto são convertidas para minúsculas e os múltiplos espaços entre as palavras são removidos. Com isso, o texto obtido é con-

vertido em uma lista de palavras.

A análise quantitativa do resultado é realizada avaliando a lista de palavras obtidas na etapa anterior com um gabarito de palavras-chave selecionadas. Nesse experimento, dois grupos de palavras-chave foram elencados. O primeiro considera os elementos necessários para rótulos de alimentos; o segundo, para rótulos de água. Assim, a distância de Levenshtein é computada entre cada palavra obtida para cada palavra-chave. Também é possível realizar a comparação de palavras compostas. Por fim, o resultado da análise consiste em uma relação numérica entre as palavras obtidas com as palavras-chave. Desse modo, busca-se, para cada palavra-chave, a palavra com o menor valor (menor distância); se esse valor estiver abaixo de um limiar pré-determinado, considera-se o termo como recuperado.

A acurácia, portanto, é computada a partir da razão entre a quantidade de termos recuperados e a quantidade de palavras-chave. São expostas duas acurácias: a maior computada individualmente e a combinada, obtida através da união das listas de termos de todas as imagens processadas. Cada imagem foi avaliada com o seu respectivo grupo de palavras-chave.

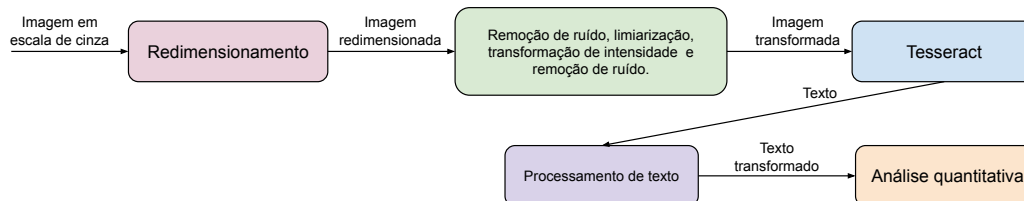


Figure 12. Fluxograma do quarto método, quinto experimento

Resultados

O resultado desse grupo de experimentos é apresentado nas tabelas 7 e 8.

A tabela 7 apresenta algumas imagens nas quais foram aplicadas as técnicas do experimento, assim como a sua respectiva imagem original. A coluna de identificação serve para guiar a visualização de resultados na tabela 8.

A tabela 8 apresenta os resultados das métricas computadas para as imagens da tabela 7. Como citado anteriormente, a maior acurácia individual, entre os processamentos realizados para cada fator de redimensionamento, é exposta. Nota-se, para os identificadores 1 e 4 uma diferença significativa em relação à acurácia combinada, que leva em consideração os termos obtidos em todos os processamentos da mesma imagem. Por exemplo, para o identificador 1, apenas "gorduras saturadas"² não foi recuperado nesse processamento individual em questão, mas, sim, em algum outro. Já no caso dos identificadores 2 e 3, o melhor resultado individual já foi capaz de recuperar todas as palavras-chave.

Dois fatores principais diferem esse grupo de experimentos com o restante. Primeiramente, nota-se, qualitativamente, o realce nos caracteres dos rótulos, promovido pela aplicação das técnicas de pré-processamento. Isso fez com que o texto obtido no

²Palavras compostas são consideradas como apenas um termo

OCR fosse mais limpo que anteriormente. Por fim, a análise quantitativa, a partir da utilização da distância de Levenstein, permitiu uma observação prática dos resultados.

Ainda com os avanços realizados nesse grupo de experimento, não foi possível obter, de forma consistente, os valores numéricos de cada palavra-chave do rótulo. A tabela 9 apresenta um exemplo³ desse fato. Nota-se, no texto de saída obtido para a primeira imagem apresentada, a falta dos valores numéricos de cada linha dos elementos. Alguns casos, como dos termos de "Valor Energético" e "Proteínas", tiveram uma recuperação de dígitos, mas de uma forma ruidosa. Entretanto, a parte inferior do rótulo, que trata da elucidação da escala de valores diários, apresentou um texto de saída preciso, inclusive com os números e unidade de medida em questão: 2000 kcal e 8400 kJ. Entretanto, para a segunda imagem, o cenário é outro. É possível verificar que a maior parte dos dígitos foram recuperados integralmente. Isso possibilita a formação de pares entre as palavras-chave com os seus respectivos valores. O conjunto combinado e de melhor resultado dessa recuperação é apresentado na tabela 10. Essa distinção de resultados acontece por algumas limitações do Tesseract, que necessita do contexto dos caracteres para um melhor reconhecimento.

3.3. Sumário Quantitativo dos Resultados

A etapa de quantificação dos resultados só foi implementada a partir do quinto grupo de experimentos. Desse modo, com intuito de apresentar métricas para os experimentos anteriores e possibilitar uma comparação formal, realizou-se mais uma etapa de experimentos. Para isso, foram elencadas cinco amostras de imagens dos conjuntos de dados A, C e D. Cada amostra de dados foi aplicada aos métodos já apresentados e, posteriormente, à etapa de processamento de texto e recuperação da métrica de acurácia. A tabela 11 apresenta os resultados obtidos, onde são elencadas a acurácia mínima, máxima e média para cada aplicação de amostra a um método. O maior resultado de cada grupo está em negrito, enquanto o segundo maior, sublinhado.

A coluna *Amostra* da tabela 11 identifica de qual conjunto de dados ela foi originada. A coluna *Método (descrição)* identifica qual foi o método aplicado às imagens daquela amostra, fazendo uma conexão com o diagrama de apresentação do método. Um método número 5 foi adicionado apenas para essa etapa; sendo considerado como uma ramificação do método 4. Nesse método, apenas o redimensionamento foi considerado na etapa de pré-processamento da imagem. As colunas seguintes apresentam as estatísticas referentes à *Acurácia* obtida em cada aplicação.

Os resultados dessa seção reforçam o que já foi apresentado anteriormente, principalmente em relação às características da imagem. É notável que os melhores resultados foram obtidos com amostras do conjunto de dados D, enquanto os piores resultados foram obtidos com a amostra do primeiro conjunto. Além disso, notamos que o último experimento (número 5) apresentou um resultado melhor, comparado com os anteriores. De fato, os melhores resultados foram com os métodos 4 e 5. Ainda assim, um dos pontos principais e de atenção desse experimento é, de fato, o redimensionamento. Isso se dá pois o método 5 foi capaz de demonstrar acurácia média maior em todos os casos, ainda que os resultados sejam bem próximos.

³ Algumas quebras de linha foram adicionadas ao texto para facilitar a visualização

3.4. Aplicação

Um protótipo de aplicação foi desenvolvido para utilização interativa do fluxo de execução proposto no experimento 3.2.5. O aplicativo `paNNe`⁴ foi desenvolvido na linguagem Python e utilizando o framework Streamlit para possibilitar a visualização e interação do usuário. Sua exibição está disponível tanto em português como em inglês.

A figura 13 apresenta o cabeçalho de apresentação do aplicativo, com seu título, subtítulo e o logo da Universidade Federal de Santa Catarina.



paNNe

Package Label Scanner

Figure 13. Cabeçalho do aplicativo

O aplicativo permite a personalização das definições disponíveis para processamento de imagem e de texto. Nesse sentido, são possíveis customizações no fluxo de pré-processamento que a imagem passará. É possível selecionar quais métodos serão utilizados, assim como quais fatores de escala devem ser aplicados no redimensionamento. Em relação ao processamento de texto, o usuário seleciona um conjunto de palavras que será utilizado na avaliação do resultado; é possível, também, remover palavras pré-definidas.

Uma imagem resultante é exposta para cada fator de escala aplicado. Com intuito de permitir uma melhor navegação e economizar espaço na tela, sua exibição só ocorre, de fato, após o usuário clicar no respectivo *drop-down*.

A seção de resultados é apresentada logo após a finalização da avaliação de cada imagem individualmente. Desse modo, são apresentadas avaliações combinadas, onde o resultado de todas as imagens avaliadas são considerados, e a melhor avaliação individual. Além da acurácia obtida, também são apresentadas as listas de termos recuperados, o dicionário de dígitos recuperados e o texto recuperado para a imagem com melhor avaliação.

4. Conclusão e Trabalhos Futuros

O reconhecimento óptico de caracteres, ainda que amplamente estudado na literatura, ainda apresenta desafios a serem ultrapassados. Diversas técnicas assumem características específicas para apresentar uma boa performance. Nesse sentido, etapas de

⁴O código fonte está disponível em <https://github.com/cchenzi/tcc>.

pré-processamento de imagens são essenciais para obter resultados que sejam úteis no cotidiano. Neste trabalho, foi apresentado uma visão geral sobre a leitura automática de rótulos de embalagens, com enfoque na utilização de um sistema de OCR de código aberto, o Tesseract. Foram realizados experimentos exploratórios, de avaliação qualitativa e quantitativa. Em ambos os cenários metodologias de pré-processamento de imagens foram abordadas e utilizadas.

Os experimentos exploratórios de avaliação qualitativa tiveram intuito de promover o entendimento sobre o comportamento do Tesseract e ampliar o conhecimento sobre técnicas de visão computacional clássica. Por se tratar de análise qualitativa, a discussão dos resultados é um tanto quanto abstrata. Notou-se que vários textos foram recuperados com termos úteis, mas que, ainda assim, deixavam a desejar em questão de consistência. Entretanto, esses experimentos foram úteis para pavimentar a trajetória para as análises quantitativas, com definições e objetivos mais concretos.

Os experimentos de avaliação quantitativa mostraram bons resultados com imagens em condições favoráveis (e.g., foco no rótulo, boa iluminação, rótulo planejado). Foi possível recuperar palavras-chave dos rótulos com uma grande precisão. Isso permite uma análise elementar dos fatos nutricionais presentes nas embalagens; torna-se, portanto, capaz de verificar as conformidades de um rótulo em relação à presença de informações básicas. Também foi apresentado um exemplo da produção de pares entre palavras-chave e seus respectivos valores numéricos. Entretanto, essa última parte ainda é muito sensível às características da imagem, uma vez que não foi possível recuperar, de forma consistente, para o restante das imagens. Em razão disso, a utilização do protótipo apresentado pode ser um pouco restrita.

Para continuação deste projeto, trabalhos futuros incluem melhorar as técnicas de pré-processamento de imagens já apresentadas, assim como explorar e desenvolver novos algoritmos que contribuam neste aspecto. Uma das possíveis melhorias consiste em realizar a extração consistente de regiões, a fim de diminuir o preparo manual que antecede à aplicação das técnicas. Além disso, também é factível o aperfeiçoamento na recuperação de caracteres numéricos, com intuito de permitir análises mais profundas em relação às necessidades e informações dos rótulos de embalagens.

References

- Agência Nacional de Vigilância Sanitária - Anvisa (2020). Rotulagem de alimentos. [Online; acessado em 08/11/2020].
- Chaudhuri, A., Mandaviya, K., Badelia, P., and Ghosh, S. K. (2017). *Optical Character Recognition Systems*, pages 9–41. Springer International Publishing, Cham.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Gonzalez, A., Bergasa, L., Yebes, J., and Bronte, S. (2012). Text location in complex images. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 617–620.

- Gundimeda, V., Murali, R., Joseph, R., and Babu, N. T. N. (2019). *An Automated Computer Vision System for Extraction of Retail Food Product Metadata: AICC 2018*, pages 199–216.
- Jain, R., Jain, R., JAIN, R., Jaina, R., Kasturi, R., Schunck, B., and Schunck, B. (1995). *Machine Vision*. Computer Science Series. McGraw-Hill.
- Khasgiwala, A. (2018). *Word recognition in nutrition labels with convolutional neural network*. PhD thesis, Utah State University.
- Kruskal, J. B. (1983). An overview of sequence comparison: Time warps, string edits, and macromolecules. *SIAM review*, 25(2):201–237.
- Matsunaga, N. and Sullivan, R. (2015). Image processing for the extraction of nutritional information from food labels.
- Nguyen, T. T. H., Jatowt, A., Nguyen, N.-V., Coustaty, M., and Doucet, A. (2020). Neural machine translation with bert for post-ocr error detection and correction. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, JCDL '20*, page 333–336, New York, NY, USA. Association for Computing Machinery.
- Patel, C., Patel, A., and Patel, D. (2012). Optical character recognition by open source ocr tool tesseract: A case study. *International Journal of Computer Applications*, 55:50–56.
- Singh, A., Bacchuwar, K., and Bhasin, A. (2012). A survey of ocr applications. *International Journal of Machine Learning and Computing (IJMLC)*.
- Singh, S. (2013). Optical character recognition techniques: a survey. *Journal of emerging Trends in Computing and information Sciences*, 4(6):545–550.
- Sporici, D., Cuşnir, E., and Boiangiu, C.-A. (2020). Improving the accuracy of tesseract 4.0 ocr engine using convolution-based preprocessing. *Symmetry*, 12:715.
- Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer-Verlag, Berlin, Heidelberg, 1st edition.
- Tesseract (2021). tessdoc - tesseract documentation. [Online; acessado em 21-abril-2021].

Identificador	Original	Processada																																																																		
1	<p>INFORMAÇÃO NUTRICIONAL Porção de 15 g (1 colher e ½ de sopa)</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor energético 26 kcal = 109 kJ</td> <td>1</td> </tr> <tr> <td>Carboidratos 0,6 g</td> <td>0</td> </tr> <tr> <td>Proteínas 0</td> <td>0</td> </tr> <tr> <td>Gorduras totais 2,6 g</td> <td>5</td> </tr> <tr> <td>Gorduras saturadas 1,6 g</td> <td>7</td> </tr> <tr> <td>Gorduras trans 0</td> <td>**</td> </tr> <tr> <td>Fibra alimentar 0</td> <td>0</td> </tr> <tr> <td>Sódio 12 mg</td> <td>1</td> </tr> </tbody> </table> <p>* % Valores Diários com base em uma dieta de 2.000 kcal ou 8.400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas. ** VD não estabelecido.</p>	Quantidade por porção	% VD (*)	Valor energético 26 kcal = 109 kJ	1	Carboidratos 0,6 g	0	Proteínas 0	0	Gorduras totais 2,6 g	5	Gorduras saturadas 1,6 g	7	Gorduras trans 0	**	Fibra alimentar 0	0	Sódio 12 mg	1	<p>▼ 0.5</p> <p>INFORMAÇÃO NUTRICIONAL Porção de 15 g (1 colher e ½ de sopa)</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor energético 26 kcal = 109 kJ</td> <td>1</td> </tr> <tr> <td>Carboidratos 0,6 g</td> <td>0</td> </tr> <tr> <td>Proteínas 0</td> <td>0</td> </tr> <tr> <td>Gorduras totais 2,6 g</td> <td>5</td> </tr> <tr> <td>Gorduras saturadas 1,6 g</td> <td>7</td> </tr> <tr> <td>Gorduras trans 0</td> <td>**</td> </tr> <tr> <td>Fibra alimentar 0</td> <td>0</td> </tr> <tr> <td>Sódio 12 mg</td> <td>1</td> </tr> </tbody> </table> <p>* % Valores Diários com base em uma dieta de 2.000 kcal ou 8.400 kJ. Seus valores diários podem ser maiores ou menores dependendo de suas necessidades energéticas. ** VD não estabelecido.</p>	Quantidade por porção	% VD (*)	Valor energético 26 kcal = 109 kJ	1	Carboidratos 0,6 g	0	Proteínas 0	0	Gorduras totais 2,6 g	5	Gorduras saturadas 1,6 g	7	Gorduras trans 0	**	Fibra alimentar 0	0	Sódio 12 mg	1																														
Quantidade por porção	% VD (*)																																																																			
Valor energético 26 kcal = 109 kJ	1																																																																			
Carboidratos 0,6 g	0																																																																			
Proteínas 0	0																																																																			
Gorduras totais 2,6 g	5																																																																			
Gorduras saturadas 1,6 g	7																																																																			
Gorduras trans 0	**																																																																			
Fibra alimentar 0	0																																																																			
Sódio 12 mg	1																																																																			
Quantidade por porção	% VD (*)																																																																			
Valor energético 26 kcal = 109 kJ	1																																																																			
Carboidratos 0,6 g	0																																																																			
Proteínas 0	0																																																																			
Gorduras totais 2,6 g	5																																																																			
Gorduras saturadas 1,6 g	7																																																																			
Gorduras trans 0	**																																																																			
Fibra alimentar 0	0																																																																			
Sódio 12 mg	1																																																																			
2	<p>Classificação: Água Mineral Fluoretada e Fracamente Radioativa na Fonte. Análise Química Boletim nº 212/LAMIN/18 de 25/07/18 LAMIN/CPRM. NÃO CONTEM GLÚTEN</p> <table border="1"> <thead> <tr> <th>Composição química (mg/L)</th> <th></th> </tr> </thead> <tbody> <tr> <td>BICARBONATO</td> <td>23,13</td> </tr> <tr> <td>SÓDIO</td> <td>2,707</td> </tr> <tr> <td>CÁLCIO</td> <td>2,849</td> </tr> <tr> <td>MAGNÉSIO</td> <td>2,350</td> </tr> <tr> <td>CLORETO</td> <td>2,150</td> </tr> <tr> <td>SULFATO</td> <td>0,690</td> </tr> <tr> <td>POTÁSSIO</td> <td>0,617</td> </tr> <tr> <td>NITRATO</td> <td>0,820</td> </tr> <tr> <td>FLUORETO</td> <td>0,040</td> </tr> <tr> <td>ESTRÔNCIO</td> <td>0,017</td> </tr> <tr> <td>BÁRIO</td> <td>0,014</td> </tr> </tbody> </table>	Composição química (mg/L)		BICARBONATO	23,13	SÓDIO	2,707	CÁLCIO	2,849	MAGNÉSIO	2,350	CLORETO	2,150	SULFATO	0,690	POTÁSSIO	0,617	NITRATO	0,820	FLUORETO	0,040	ESTRÔNCIO	0,017	BÁRIO	0,014	<p>▼ 0.5</p> <p>Classificação: Água Mineral Fluoretada e Fracamente Radioativa na Fonte. Análise Química Boletim nº 212/LAMIN/18 de 25/07/18 LAMIN/CPRM. NÃO CONTEM GLÚTEN</p> <table border="1"> <thead> <tr> <th>Composição química (mg/L)</th> <th></th> </tr> </thead> <tbody> <tr> <td>BICARBONATO</td> <td>23,13</td> </tr> <tr> <td>SÓDIO</td> <td>2,707</td> </tr> <tr> <td>CÁLCIO</td> <td>2,849</td> </tr> <tr> <td>MAGNÉSIO</td> <td>2,350</td> </tr> <tr> <td>CLORETO</td> <td>2,150</td> </tr> <tr> <td>SULFATO</td> <td>0,690</td> </tr> <tr> <td>POTÁSSIO</td> <td>0,617</td> </tr> <tr> <td>NITRATO</td> <td>0,820</td> </tr> <tr> <td>FLUORETO</td> <td>0,040</td> </tr> <tr> <td>ESTRÔNCIO</td> <td>0,017</td> </tr> <tr> <td>BÁRIO</td> <td>0,014</td> </tr> </tbody> </table>	Composição química (mg/L)		BICARBONATO	23,13	SÓDIO	2,707	CÁLCIO	2,849	MAGNÉSIO	2,350	CLORETO	2,150	SULFATO	0,690	POTÁSSIO	0,617	NITRATO	0,820	FLUORETO	0,040	ESTRÔNCIO	0,017	BÁRIO	0,014																		
Composição química (mg/L)																																																																				
BICARBONATO	23,13																																																																			
SÓDIO	2,707																																																																			
CÁLCIO	2,849																																																																			
MAGNÉSIO	2,350																																																																			
CLORETO	2,150																																																																			
SULFATO	0,690																																																																			
POTÁSSIO	0,617																																																																			
NITRATO	0,820																																																																			
FLUORETO	0,040																																																																			
ESTRÔNCIO	0,017																																																																			
BÁRIO	0,014																																																																			
Composição química (mg/L)																																																																				
BICARBONATO	23,13																																																																			
SÓDIO	2,707																																																																			
CÁLCIO	2,849																																																																			
MAGNÉSIO	2,350																																																																			
CLORETO	2,150																																																																			
SULFATO	0,690																																																																			
POTÁSSIO	0,617																																																																			
NITRATO	0,820																																																																			
FLUORETO	0,040																																																																			
ESTRÔNCIO	0,017																																																																			
BÁRIO	0,014																																																																			
3	<p>INGREDIENTES: leite integral esterilizado; refinado de açúcar (NO 421), citrato de cálcio (NO 2318), mesofosfato monobásico (NO 1219) e dióxido de cálcio (NO 430).</p> <p>ALÉRGICOS: CONTEM LACTE. CONTEM LACTOSE. NÃO CONTEM GLÚTEN.</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor Energético 116 kcal = 487 kJ</td> <td>6</td> </tr> <tr> <td>Carboidratos 9,1 g</td> <td>3</td> </tr> <tr> <td>Proteínas 6,3 g</td> <td>8</td> </tr> <tr> <td>Gorduras Totais 6,0 g</td> <td>11</td> </tr> <tr> <td>Gorduras Saturadas 4,3 g</td> <td>20</td> </tr> <tr> <td>Gorduras Trans 0 g</td> <td>**</td> </tr> <tr> <td>Fibra Alimentar 0 g</td> <td>0</td> </tr> <tr> <td>Sódio 152 mg</td> <td>6</td> </tr> <tr> <td>Calcio 209 mg</td> <td>21</td> </tr> </tbody> </table> <p>* % Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** Valor não especificado.</p> <p>MANTENHA EM LOCAL SECO E AREJADO. APÓS ABERTURA, CONSERVAR SOB REFRIGERAÇÃO (2-8 °C). NÃO É NECESSÁRIO FEVERIL.</p>	Quantidade por porção	% VD (*)	Valor Energético 116 kcal = 487 kJ	6	Carboidratos 9,1 g	3	Proteínas 6,3 g	8	Gorduras Totais 6,0 g	11	Gorduras Saturadas 4,3 g	20	Gorduras Trans 0 g	**	Fibra Alimentar 0 g	0	Sódio 152 mg	6	Calcio 209 mg	21	<p>▼ 0.9</p> <table border="1"> <thead> <tr> <th>Quantidade por porção</th> <th>% VD (*)</th> </tr> </thead> <tbody> <tr> <td>Valor Energético 116 kcal = 487 kJ</td> <td>6</td> </tr> <tr> <td>Carboidratos 9,1 g</td> <td>3</td> </tr> <tr> <td>Proteínas 6,3 g</td> <td>8</td> </tr> <tr> <td>Gorduras Totais 6,0 g</td> <td>11</td> </tr> <tr> <td>Gorduras Saturadas 4,3 g</td> <td>20</td> </tr> <tr> <td>Gorduras Trans 0 g</td> <td>**</td> </tr> <tr> <td>Fibra Alimentar 0 g</td> <td>0</td> </tr> <tr> <td>Sódio 152 mg</td> <td>6</td> </tr> <tr> <td>Calcio 209 mg</td> <td>21</td> </tr> </tbody> </table> <p>* % Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** Valor não especificado.</p>	Quantidade por porção	% VD (*)	Valor Energético 116 kcal = 487 kJ	6	Carboidratos 9,1 g	3	Proteínas 6,3 g	8	Gorduras Totais 6,0 g	11	Gorduras Saturadas 4,3 g	20	Gorduras Trans 0 g	**	Fibra Alimentar 0 g	0	Sódio 152 mg	6	Calcio 209 mg	21																										
Quantidade por porção	% VD (*)																																																																			
Valor Energético 116 kcal = 487 kJ	6																																																																			
Carboidratos 9,1 g	3																																																																			
Proteínas 6,3 g	8																																																																			
Gorduras Totais 6,0 g	11																																																																			
Gorduras Saturadas 4,3 g	20																																																																			
Gorduras Trans 0 g	**																																																																			
Fibra Alimentar 0 g	0																																																																			
Sódio 152 mg	6																																																																			
Calcio 209 mg	21																																																																			
Quantidade por porção	% VD (*)																																																																			
Valor Energético 116 kcal = 487 kJ	6																																																																			
Carboidratos 9,1 g	3																																																																			
Proteínas 6,3 g	8																																																																			
Gorduras Totais 6,0 g	11																																																																			
Gorduras Saturadas 4,3 g	20																																																																			
Gorduras Trans 0 g	**																																																																			
Fibra Alimentar 0 g	0																																																																			
Sódio 152 mg	6																																																																			
Calcio 209 mg	21																																																																			
4	<table border="1"> <thead> <tr> <th colspan="2">TABELA NUTRICIONAL</th> <th></th> </tr> <tr> <th colspan="2">Porção de 100 g (1 porção)</th> <th></th> </tr> <tr> <th>Quantidade por porção</th> <th></th> <th>% VD (**)</th> </tr> </thead> <tbody> <tr> <td>Valor Energético</td> <td>3Kcal / 12kJ</td> <td>0</td> </tr> <tr> <td>Carboidratos</td> <td>2,3 g</td> <td>1</td> </tr> <tr> <td>Proteínas</td> <td>0,3 g</td> <td>0</td> </tr> <tr> <td>Gorduras Totais</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Gorduras Saturadas</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Gorduras Trans</td> <td>0 g</td> <td>**</td> </tr> <tr> <td>Fibra Alimentar</td> <td>1,9 g</td> <td>7</td> </tr> <tr> <td>Sódio</td> <td>0,3 mg</td> <td>0</td> </tr> </tbody> </table>	TABELA NUTRICIONAL			Porção de 100 g (1 porção)			Quantidade por porção		% VD (**)	Valor Energético	3Kcal / 12kJ	0	Carboidratos	2,3 g	1	Proteínas	0,3 g	0	Gorduras Totais	0 g	0	Gorduras Saturadas	0 g	0	Gorduras Trans	0 g	**	Fibra Alimentar	1,9 g	7	Sódio	0,3 mg	0	<p>▼ 0.3</p> <table border="1"> <thead> <tr> <th colspan="2">TABELA NUTRICIONAL</th> <th></th> </tr> <tr> <th colspan="2">Porção de 100 g (1 porção)</th> <th></th> </tr> <tr> <th>Quantidade por porção</th> <th></th> <th>% VD (**)</th> </tr> </thead> <tbody> <tr> <td>Valor Energético</td> <td>3Kcal / 12kJ</td> <td>0</td> </tr> <tr> <td>Carboidratos</td> <td>2,3 g</td> <td>1</td> </tr> <tr> <td>Proteínas</td> <td>0,3 g</td> <td>0</td> </tr> <tr> <td>Gorduras Totais</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Gorduras Saturadas</td> <td>0 g</td> <td>0</td> </tr> <tr> <td>Gorduras Trans</td> <td>0 g</td> <td>**</td> </tr> <tr> <td>Fibra Alimentar</td> <td>1,9 g</td> <td>7</td> </tr> <tr> <td>Sódio</td> <td>0,3 mg</td> <td>0</td> </tr> </tbody> </table>	TABELA NUTRICIONAL			Porção de 100 g (1 porção)			Quantidade por porção		% VD (**)	Valor Energético	3Kcal / 12kJ	0	Carboidratos	2,3 g	1	Proteínas	0,3 g	0	Gorduras Totais	0 g	0	Gorduras Saturadas	0 g	0	Gorduras Trans	0 g	**	Fibra Alimentar	1,9 g	7	Sódio	0,3 mg	0
TABELA NUTRICIONAL																																																																				
Porção de 100 g (1 porção)																																																																				
Quantidade por porção		% VD (**)																																																																		
Valor Energético	3Kcal / 12kJ	0																																																																		
Carboidratos	2,3 g	1																																																																		
Proteínas	0,3 g	0																																																																		
Gorduras Totais	0 g	0																																																																		
Gorduras Saturadas	0 g	0																																																																		
Gorduras Trans	0 g	**																																																																		
Fibra Alimentar	1,9 g	7																																																																		
Sódio	0,3 mg	0																																																																		
TABELA NUTRICIONAL																																																																				
Porção de 100 g (1 porção)																																																																				
Quantidade por porção		% VD (**)																																																																		
Valor Energético	3Kcal / 12kJ	0																																																																		
Carboidratos	2,3 g	1																																																																		
Proteínas	0,3 g	0																																																																		
Gorduras Totais	0 g	0																																																																		
Gorduras Saturadas	0 g	0																																																																		
Gorduras Trans	0 g	**																																																																		
Fibra Alimentar	1,9 g	7																																																																		
Sódio	0,3 mg	0																																																																		

Table 7. Resultados obtidos no quinto grupo de experimentos

Identificador	Melhor Acurácia Individual	Termos recuperados (individual)	Acurácia Combinada	Termos recuperados (combinado)
1	0.87	"carboidratos", "proteinas", "gorduras totais", "gorduras trans", "fibra alimentar", "sodio", "valor energetico"	1	"valor energetico", "carboidratos", "gorduras saturadas", "proteinas", "gorduras totais", "fibra alimentar", "sodio", "gorduras trans"
2	1	"fluoreto", "bario", "sodio", "sulfato", "cloreto", "magnesio", "bicarbonato", "nitrato", "calcio", "estroncio", "potassio"	1	"fluoreto", "bario", "sodio", "sulfato", cloreto", "magnesio", "bicarbonato", "nitrato", "calcio", "estroncio", "potassio"
3	1	"valor energetico", "carboidratos", "gorduras saturadas", "proteinas", "gorduras totais", "fibra alimentar", "sodio", "gorduras trans"	1	"valor energetico", "carboidratos", "gorduras saturadas", "proteinas", "gorduras totais", "fibra alimentar", "sodio", "gorduras trans"
4	0.5	"proteinas", "gorduras totais", "gorduras saturadas", "gorduras trans"	0.62	"gorduras saturadas", "proteinas", "gorduras totais", "fibra alimentar", "gorduras trans"

Table 8. Resultados obtidos no quinto grupo de experimentos

Imagem

▼ 0.9

INFORMAÇÃO NUTRICIONAL		
Porção de 200 ml (1 copo)		
	Quantidade por porção	% VD (*)
Valor Energético	116 kcal = 487 kJ	6
Carboidratos	9,1 g	3
Proteínas	6,3 g	8
Gorduras Totais	6,0 g	11
Gorduras Saturadas	4,3 g	20
Gorduras Trans	0 g	**
Fibra Alimentar	0 g	0
Sódio	152 mg	6
Cálcio	209 mg	21

* % Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** Valor não especificado.

Resultado Obtido

INFORMAÇÃO NUTRICIONAL

Porção de 200 ml (1 copo) Quantidade por porção
 Valor Energético — 116kcal=487kJ — 6 8 Carboidratos Proteínas 63g — 8

Gorduras Totais

Gorduras Saturadas

Gorduras Trans Fibra Alimentar

ódio Cálcio

% Valores Diários de referência com base em uma dieta de 2000 kcal ou 8400 kJ. Seus valores diários podem ser maiores ou menores, dependendo de suas necessidades energéticas. ** Valor não especificado.

▼ 0.5

Classificação: Água Mineral Fluoretada e Fracamente Radioativa na Fonte. Análise Química Boletim nº 212/LAMIN/18 de 25/07/18 LAMIN/CPRM.

NÃO CONTÉM GLUTEN

Composição química (mg L)	
BICARBONATO	23,13
SÓDIO	2,707
CÁLCIO	2,649
MAGNÉSIO	2,350
CLORETO	2,150
SULFATO	0,690
POTÁSSIO	0,617
NITRATO	0,820
FLUORETO	0,648
EXTRÔNIO	0,017
DÁRIO	0,014

Classificação: Água Mineral Fluoretada é Fracamente Radioativa na Fonte. Análise Química Boletim nº 212/ L AMIN/18 de 25/07/18 LAMIN/CPRM. NÃO CONTÉM GLUTEN BICARBONATO 23,13 SÓDIO 2,707 CÁLCIO 2,649 MAGNÉSIO 2,350 CLORETO 2,150 SULFATO 0,690 POTÁSSIO 0617 MITRATO 0,820 FLUORETO 0,648 EXTRÔNIO e DÁRIO 01

Table 9. Resultados obtido no quinto grupo de experimentos

Identificador	Melhor Resultado	Resultado Combinado
2	"bicarbonato": "23,13", "sodio": "2,707", "calcio": "2,649", "magnesio": "2,350", "cloreto": "2,150", "sulfato": "0,690", "potassio": "0617", "nitrato": "0,820", "fluoreto": "0,648", "bario": "01"	"bicarbonato": ["23,13"], "sodio": ["2,707"], "calcio": ["2,849", "2,649"], "magnesio": ["2,350"], "cloreto": ["2,150"], "sulfato": ["0,690"], "potassio": ["0617", "0,617"], "nitrato": ["0,820"], "fluoreto": ["0,648", "0,640"], "bario": ["0,014", "01"]

Table 10. Exemplo de recuperação de dígitos no quinto grupo de experimentos

Amostra	Método (descrição)	Acurácia Mínima	Acurácia Máxima	Acurácia Média
1	1 (fig. 8)	0	0.27	0.12
1	2 (fig. 9)	0	0	0
1	4 (fig. 12)	0	<u>0.54</u>	<u>0.25</u>
1	5 (fig. 12) *	0	0.63	0.29
3	1 (fig. 8)	0	0.27	0.07
3	2 (fig. 9)	0	<u>0.36</u>	0.07
3	4 fig. 12)	0.09	1	<u>0.52</u>
3	5 (fig. 12) *	0.09	1	0.56
4	1 (fig. 8)	0	1	0.3
4	2 (fig. 9)	0	<u>0.75</u>	0.29
4	4 fig. 12)	<u>0.81</u>	1	<u>0.93</u>
4	5 (fig. 12) *	0.87	1	0.95

Table 11. Sumário quantitativo para comparação dos métodos apresentados.

* A etapa de transformação da imagem não foi realizada.

Em negrito, destaca-se a melhor acurácia para aquele grupo; sublinhada, a segunda melhor acurácia.