



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

A ficha de identificação é elaborada pelo próprio autor.

Orientações em:

<http://portalbu.ufsc.br/ficha>

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, 07 de maio de 2021.

Prof. XXXXXX, Dr.
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

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, apresentaremos neste projeto 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, valores nutricionais, etc. 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, apresentaremos uma revisão bibliográfica do Estado da Arte e métricas para avaliar o desempenho do modelo utilizado.

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

LISTA DE FIGURAS

Figura 1 – Foto de um cachorro dormindo	12
Figura 2 – Representação matricial resumida da imagem à esquerda	12
Figura 3 – Exemplo de aplicação de limiarização	13
Figura 4 – Exemplo de dilatação e erosão	14
Figura 5 – Exemplo de abertura	14
Figura 6 – Exemplo de fechamento	15
Figura 7 – Rótulos com ângulo original	17
Figura 8 – Rótulo com ângulo original, aproximado	17
Figura 9 – Rótulo planificado, sobre superfície plana	18
Figura 10 – Componentes extraídos de um rótulo planificado	18
Figura 11 – Fluxograma do primeiro experimento	19
Figura 12 – Exemplo de imagem que não produziu resultado no primeiro grupo dos experimentos	20
Figura 13 – 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	21
Figura 14 – Fluxograma do segundo experimento	21
Figura 15 – 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.	22
Figura 16 – Fluxograma do quarto experimento	24
Figura 17 – 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.	26

LISTA DE TABELAS

Tabela 1 – Sumário do conjunto de dados	19
Tabela 2 – Resultado obtido no primeiro grupo de experimentos.	20
Tabela 3 – Resultado de uma imagem do segundo experimento, dividido pelas técnicas aplicadas	23
Tabela 4 – Resultados obtido no terceiro grupo de experimentos	25
Tabela 5 – Resultados obtido no quarto grupo de experimentos	27

SUMÁRIO

1	INTRODUÇÃO	8
1.1	OBJETIVOS	9
1.2	OBJETIVOS ESPECÍFICOS	9
1.3	ESTRUTURA DO TRABALHO	9
2	CONCEITOS BÁSICOS E TRABALHOS RELACIONADOS	11
2.1	CONCEITOS BÁSICOS	11
2.1.1	Aprendizado de Máquina	11
2.1.2	Visão Computacional	11
2.1.2.1	Limiarização	12
2.1.2.2	Operadores morfológicos	13
2.1.2.3	Análise de regiões	14
2.1.3	Reconhecimento Ótico de Caracteres	15
2.1.3.1	Tesseract	15
2.2	TRABALHOS RELACIONADOS	15
3	DESENVOLVIMENTO	17
3.1	CONJUNTO DE DADOS	17
3.2	EXPERIMENTOS	19
3.2.1	Primeiro experimento	19
3.2.1.1	Descrição	19
3.2.1.2	Resultados	19
3.2.2	Segundo experimento	21
3.2.2.1	Descrição	21
3.2.2.2	Resultados	22
3.2.3	Terceiro experimento	22
3.2.3.1	Descrição	22
3.2.3.2	Resultados	24
3.2.4	Quarto experimento	24
3.2.4.1	Descrição	24
3.2.4.2	Resultados	24
4	CONCLUSÃO	28
	REFERÊNCIAS	29

1 INTRODUÇÃO

Rótulos correspondem, em sua maioria, à fonte primária de informação a cerca 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 definidos, 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 regulatória. 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 regulatória 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 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 a 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, utilizaremos 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 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. disponibilizar código-fonte e 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 expostos 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 feito 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 apresentado 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 estimativa de tempo de estadia em hospitais, e problemas de classificação, como para predizer 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 desenvolvemos 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 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 à 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 pixeis podem assumir apenas dois valores, 0 e 1.



Figura 1 – Foto de um cachorro dormindo

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

Figura 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.

2.1.2.1 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 aplicada limiarização, a fim de representá-la de forma binária. Neste exemplo, foi utilizado um limiar de 75, selecionado arbitrariamente. Entretanto, não necessariamente este é o valor ótimo para realizar a limiarização. Existem técnicas de obter automaticamente os valores de limiar, onde as regiões são analisadas a fim de identificar resultados mais acurados.

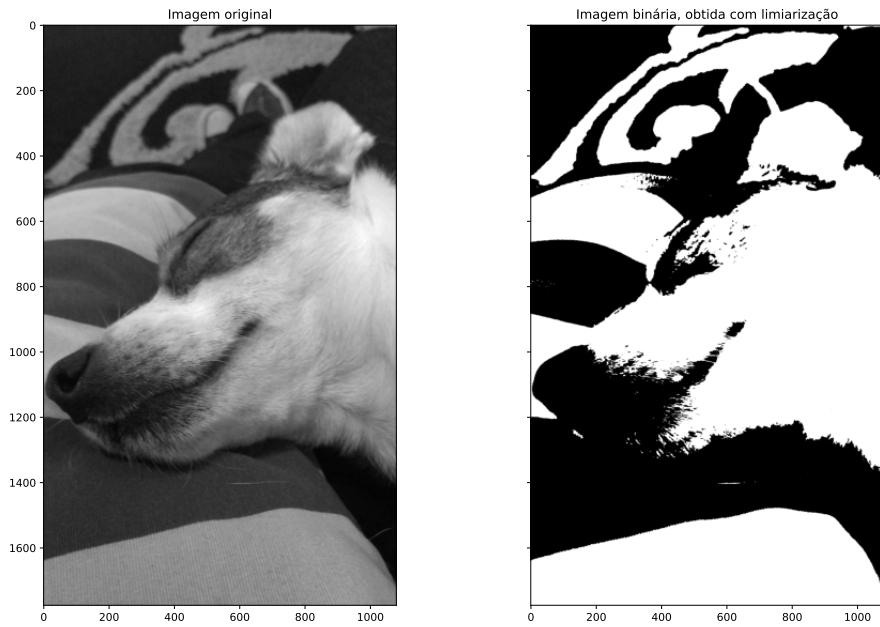


Figura 3 – Exemplo de aplicação de limiarização

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 de convolução 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

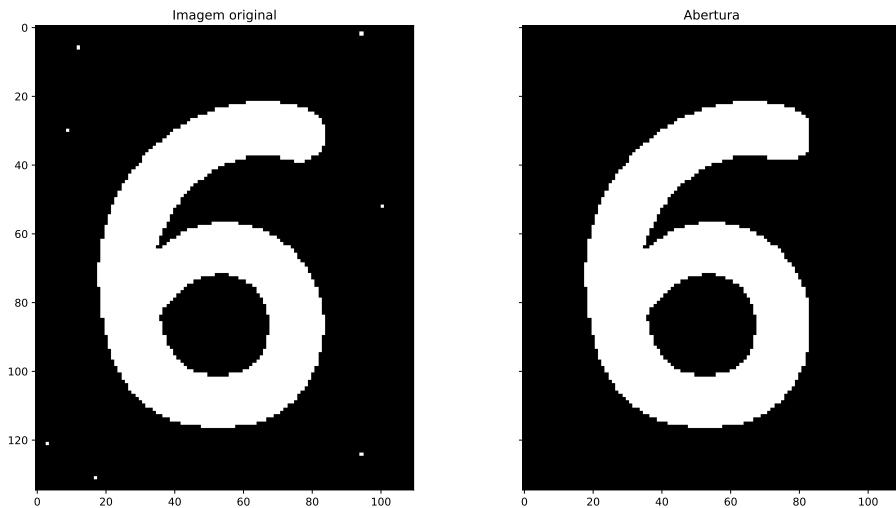


Figura 5 – Exemplo de abertura

2.1.2.3 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). Exemplos desses métodos variam desde a aplicação de divisão e junção de regiões até a utilização de histogramas de intensidade.

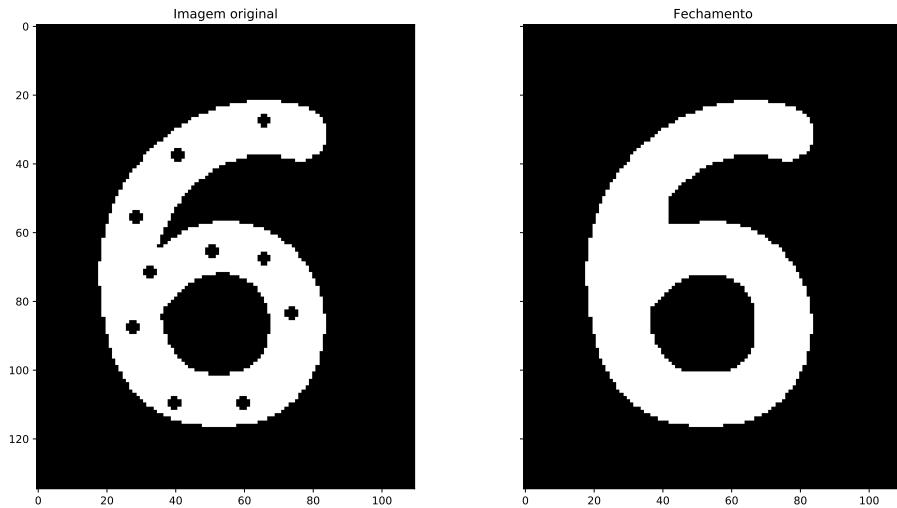


Figura 6 – Exemplo de fechamento

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 aberto¹, 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 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.

2.2 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

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

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 adaptativas.

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 Boiangiu (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*.

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. Em busca de uma maior diversidade, foram selecionadas quatro marcas diferentes. Ainda assim, as imagens foram separadas em três grupos para os experimentos, dados características como resolução e ângulo das fotos.

O primeiro grupo 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 7 mostra três exemplos de imagens para o primeiro caso, enquanto a figura 8 apresenta um exemplo para o segundo.



Figura 7 – Rótulos com ângulo original



Figura 8 – Rótulo com ângulo original, aproximado

Para o segundo grupo 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 9 é possível observar um exemplo deste grupo.



Figura 9 – Rótulo planificado, sobre superfície plana



Figura 10 – Componentes extraídos de um rótulo planificado

Para o terceiro grupo do conjunto de dados, foram extraídos componentes menores a partir das figuras do segundo conjunto de dados. A figura 10 apresenta dois componentes recortados da figura 9.

O sumário do conjunto de dados é apresentado na tabela 1, dividido pelos grupos, junto às suas características previamente discutidas e suas respectivas quantidades.

Grupo	Características	Quantidade
1	Corpo inteiro da garrafa, ângulo e formato original mantidos, resolução grande	44
2	Rótulos planificados, grande resolução	26
3	Recortes de elementos dos rótulos planificados, baixa resolução	13

Tabela 1 – 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 quatro 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 primeiro grupo do conjunto de dados. A ideia principal consistiu em observar o que teríamos como resposta neste caso, conferir como seria a performance do sistema com as características apresentadas pelas imagens. A figura 11 apresenta um fluxograma que sumariza o primeiro experimento.

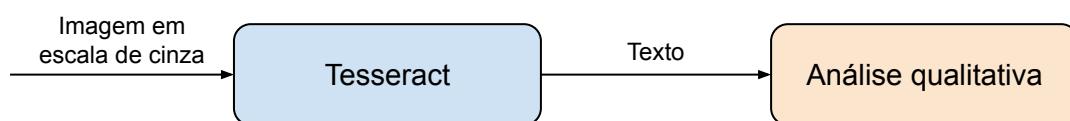


Figura 11 – Fluxograma do primeiro experimento

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 12. 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 13 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 2.

Imagen	Resultado
Original, em escala de cinza	WOCONTEM (COMPOSIGAO (CARACTERGTERY GLUTEN. QUIMICA (mg/L) FISICO-ULMICAs 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 2 – Resultado obtido no primeiro grupo de experimentos.



Figura 12 – Exemplo de imagem que não produziu resultado no primeiro grupo dos experimentos

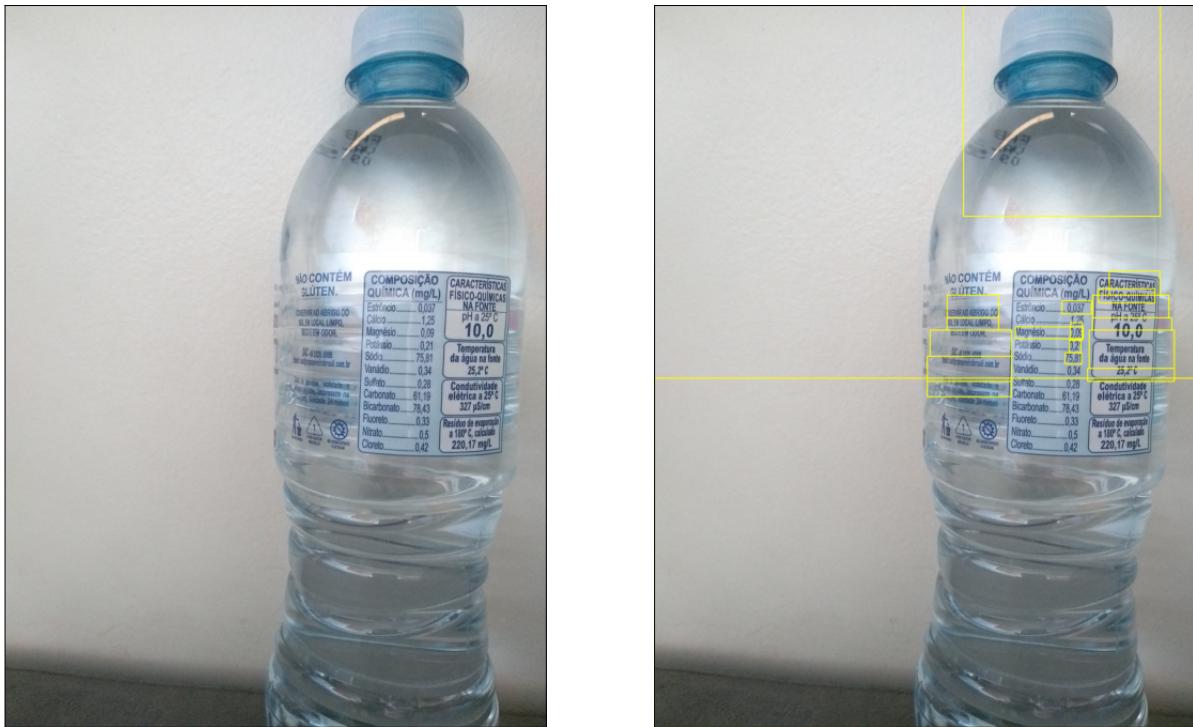


Figura 13 – Exemplo obtido no primeiro grupo dos experimentos. À esquerda, a imagem original. À direita, a imagem com as caixas de identificação obtidas com o Tesseract

3.2.2 Segundo experimento

3.2.2.1 Descrição

O segundo grupo de experimentos também foi realizado com o primeiro conjunto 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 barulho mediante borrão, limiarização e Canny.

Uma série de experimentos foi realizada, combinando as técnicas e experimentando alguns parâmetros. As etapas mencionadas são apresentadas na figura 14.

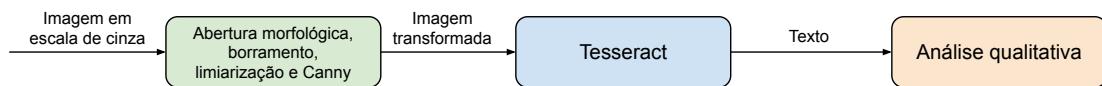


Figura 14 – Fluxograma do segundo experimento

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 o borramento, 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 tantos resultados bons neste experimento. Entretanto, a figura 15 apresenta um exemplo que produziu resultados interessantes. Vale notar que a imagem original é a mesma da figura 13. Os resultados para esse exemplo estão apresentados na tabela 3. Notamos 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.

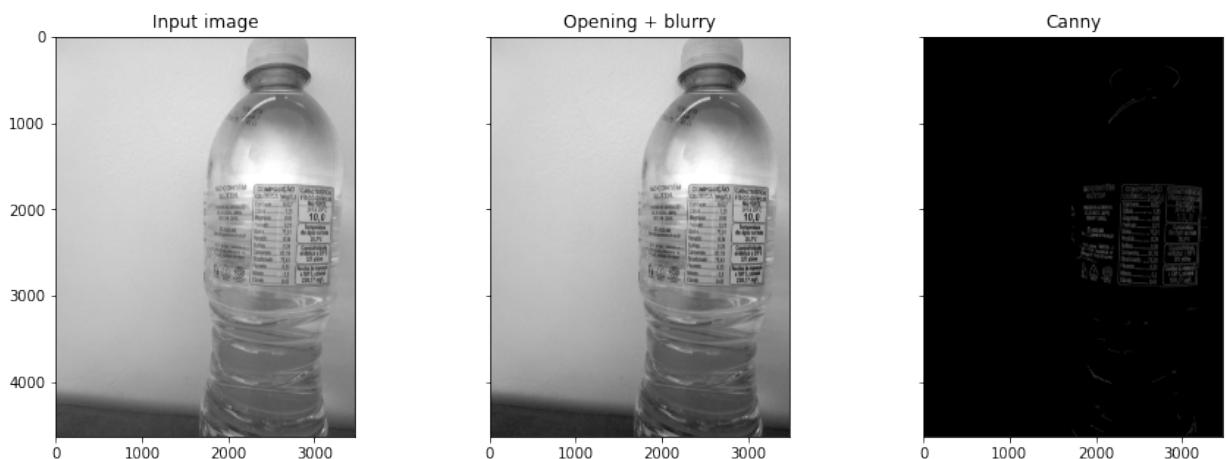


Figura 15 – Exemplo obtido no segundo grupo dos experimentos. À esquerda, a imagem original, em escala de cinza. No meio, a imagem aplicada à abertura e embacamento. À direita, a imagem aplicada ao Canny.

3.2.3 Terceiro experimento

3.2.3.1 Descrição

O terceiro grupo de experimentos foi realizado com o terceiro grupo do conjunto de dados. Em relação à metodologia aplicada, este experimento assemelha-se com o primeiro (fluxograma 11, 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.

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

Tabela 3 – Resultado de uma imagem do segundo experimento, dividido pelas técnicas aplicadas

3.2.3.2 Resultados

A tabela 4 apresenta três exemplos de resultados obtidos. Para a primeira imagem, observamos 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, verificamos que a informação recuperada não foi tão boa. Notamos 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 e último experimento, 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 16, enquanto a figura 17 apresenta um exemplo prático destas etapas.

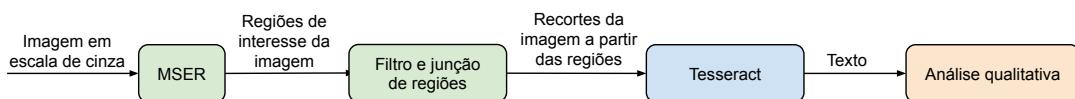


Figura 16 – Fluxograma do quarto experimento

3.2.4.2 Resultados

Um exemplo de resultado deste quarto grupo de experimentos é apresentado na tabela 5. Entretanto, observamos que os resultados não foram tão satisfatórios, tanto na

Imagen	Resultado Obtido
	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 Paraná - 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.
	Classificagio: Agua Mineral Fluoretada e Fracaimente Radioativa na Fonte. Andlse Quimica Boletim n° 212/LAMIN/18 de 25/07/18 LAMIN/CPRM, NAO CONTEM GLUTEN Composicao quimica (mg/L) BICARBONATO 2313 s0DI0 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
	NAO CONTEM GLUTEN. CONSERVAR AO ABRIGO DO SOL, EM LOCAL LIMPO, SECO E SEM ODOR. Temperatura SAC-45125.5000 da agua na fonte Emil: sac@grupopereirabrasil com 25,2°C ? i Condutividade Data de envase, validade e . : h elétric和平 25°C Wmfig, "m.mvagd;"dem i 327 pS/cm

Tabela 4 – Resultados obtido no terceiro grupo de experimentos

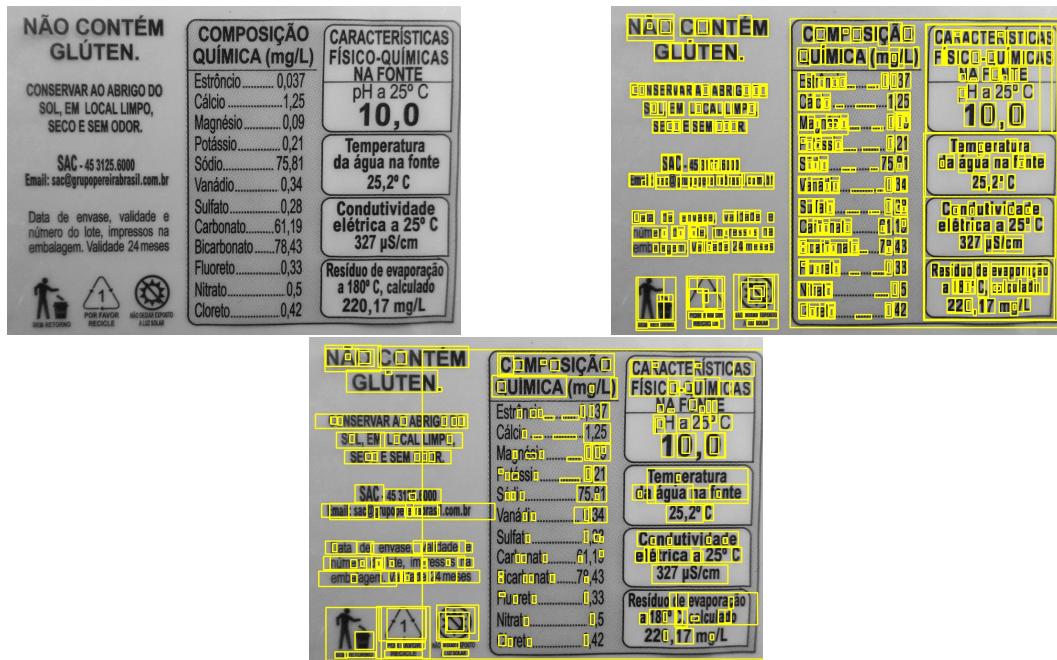


Figura 17 – Exemplo de processo do experimento quatro. Na esquerda superior, a imagem original em escala de cinza. Na direita superior, a imagem após MSER e filtragem de regiões. Na parte inferior centralizada, a imagem após junção das regiões.

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.

Imagen	Resultado Obtido
	<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 25° C 327 uS/cm Resíduo de evaporação 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>

Tabela 5 – Resultados obtido no quarto grupo de experimentos

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 com intuito de entender o comportamento deste sistema, onde metodologias de pré-processamento de imagens foram abordadas e utilizadas.

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. Além disso, também serão feitas análises quantitativas, apresentando métricas formais de desempenho dos métodos estudados.

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>.

JAIN, R. *et al.* **Machine Vision**. [S.l.]: McGraw-Hill, 1995. (Computer Science Series). ISBN 9780070320185.

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; CUŞNIR, 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/>.