UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO

Thais de Oliveira Caldeira

Sistema de Reconhecimento Óptico de Caracteres Aplicado ao Controle de Qualidade da Identificação de Bobinas de Aço

Vitória - ES

Dezembro/2018

Thais de Oliveira Caldeira

Sistema de Reconhecimento Óptico de Caracteres Aplicado ao Controle de Qualidade da Identificação de Bobinas de Aço

Parte manuscrita do Projeto de Graduação da aluna Thais de Oliveira Caldeira, apresentada ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheira Eletricista.

Universidade Federal do Espírito Santo
Centro Tecnológico
Departamento de Engenharia Elétrica

Vitória - ES
Dezembro/2018

Thais de Oliveira Caldeira

Sistema de Reconhecimento Óptico de Caracteres Aplicado ao Controle de Qualidade da Identificação de Bobinas de Aço

Parte manuscrita do Projeto de Graduação da aluna Thais de Oliveira Caldeira, apresentada ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheira Eletricista.

Aprovado em 13 de Dezembro de 2018.

COMISSÃO EXAMINADORA:

Prof. Dr. Patrick Marques Ciarelli
Universidade Federal do Espírito Santo
Orientador

Eng. Gentil Auer Neto
ArcelorMittal Tubarão
Coorientador

Prof. Dr. Klaus Fabian Côco
Universidade Federal do Espírito Santo
Examinador

Msc. Bruno Légora Souza da Silva
Universidade Federal do Espírito Santo
Examinador

Vitória - ES

Dezembro 2018

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador, Patrick, sua gigantesca paciência, imensa disposição, seus ótimos pitacos e sua inúmeras revisões ao longe do desenvolvimento deste trabalho. Obrigada principalmente pela atenção.

À equipe da Automação Nível 2 do LTQ da ArcelorMittal Tubarão, que possibilitou este trabalho e também me ofereceu um ano de muito aprendizado, – bolos e mentiras – e boa companhia.

À Universidade Federal do Espírito Santo, seus professores e funcionários, que ao longo deste curso me criou na diversidade, me oferecendo acesso à diferentes mundos e principalmente uma educação de qualidade.

À minha família, meus amigos, à Karin, ao Aleatórolho, eu amo vocês.

Por fim, e não menos importante, à banca examinadora pela aceitação do convite e pelo tempo investido para leitura e avaliação desse trabalho

por um lindésimo de segundo

tudo em mim
anda a mil
tudo assim
tudo por um fio
tudo feito
tudo estivesse no cio
tudo pisando macio
tudo psiu

tudo em minha volta
anda às tontas
como se as coisas
fossem todas
afinal de contas

(Paulo Leminski)

RESUMO

Os novos desafios da indústria moderna, gerados em torno da exigência por uma produção com mais qualidade e economia vêm exigindo o uso de tecnologia afim de otimizar os processos. Neste cenário, a chamada manufatura inteligente vem sendo usada e desenvolvida. É considerada como um novo modelo de fabricação que propõe soluções inteligentes no intuito de aprimorar o projeto, produção, gerenciamento e integração de todo o ciclo de vida de um produto típico. Por exemplo, o gerenciamento para garantir a entrega dos produtos corretos aos clientes pode aumentar a confiabilidade entre as partes e evitar multas e desperdícios desnecessários. Nesse contexto, no presente trabalho é apresentado um sistema projetado para detectar erros de impressão de identificação em bobinas de aço que possam causar problemas de rastreamento ou culminar na entrega do produto errado ao cliente final.

Propõe-se o uso de um sistema de reconhecimento óptico de caracteres para extrair a identificação impressa de bobinas de aço a partir de imagens capturadas por uma câmera fixa em um ambiente industrial. O método considera diferentes técnicas de processamento digital de imagens para lidar com a significativa variação de iluminação e impressão observada, seguida por um processo de segmentação que extrai e alinha os caracteres originalmente impressos em forma de arco, finalizando com uma rotina de classificação baseada em uma rede neural convolucional. Experimentos realizados em um conjunto de dados com aproximadamente 20.000 imagens alcançou uma acurácia média maior que 98%, suportando a validade do método proposto.

Palavras-chave: Reconhecimento Óptico de Caracteres; Manufatura Inteligente; Processamento Digital de Imagens; Redes Neurais Convolucionais.

ABSTRACT

The new challenges of modern industry, often caused by the high demand for a production with more quality and economy, require the use of technology in order to optimize the processes. In this scenario, the so-called intelligent manufacturing has been used and developed, considered as a new manufacturing model that proposes intelligent solutions in order to improve the design, production, management and integration of the entire life cycle of a typical product. For example, management to ensure delivery of the right products to customers can increase reliability between parties and avoid unnecessary fines and waste. In this context, in this work is presented a system designed to detect identification printing errors in steel coils that can cause tracking problems or lead to the delivery of the wrong product to the final customer.

It is proposed to use an optical character recognition system to extract the printed identification of steel coils from images captured by a fixed camera in an industrial environment. The method considers different techniques of digital image processing to deal with the significant variation of illumination and impression observed, followed by a segmentation process that extracts and aligns the characters originally printed in an arc form, ending with a classification routine based on a convolutional neural network. Experiments performed on a dataset with approximately 20,000 images reached an average accuracy greater than 98%, supporting the validity of the proposed method.

Keywords: Optical Character Recognition; Intelligent manufacturing; Digital Image Processing; Convolutional Neural Networks.

LISTA DE FIGURAS

Figura 1 – Exemplo de bobina de aço com identificação impressa. 20
Figura 2 – Exemplo de uso do filtro morfológico *Top Hat* 25
Figura 3 – Seção transversal de um filtro homomórfico circular simétrico no domínio da frequência 26
Figura 4 – Tipos de transformações afins 29
Figura 5 – Ilustração da operação de convolução 2D 30
Figura 6 – Ilustração da operação de *max-pooling* 31
Figura 7 – Fluxograma do Sistema Proposto 33
Figura 8 – Seleção da região de interesse. 34
Figura 9 – Comparação do uso dos filtros após a binarização de Otsu mostrando em cada linha um diferente tipo de imagem encontrada. O filtro *Top Hat* filtrou adequadamente apenas a imagem da linha C, enquanto todos os outros exemplos necessitaram dos filtros B e C homomórficos. 36
Figura 10 – Exemplos de utilização da binarização adaptativa. 37
Figura 11 – Transformação geométrica para compensar deformação devido à aquisição oblíqua: À esquerda, imagem binarizada, à direita, imagem após a transformação e o preenchimento na base. 39
Figura 12 – Da esquerda para a direita: Imagem original; imagem filtrada pelo Filtro B (homomórfico) e binarizada; imagem redimensionada e sem pequenos componentes conectados. 40
Figura 13 – Retângulos inscritos alinhados à referência 41
Figura 14 – Imagem final dos caracteres alinhados 42
Figura 15 – Exemplos de imagens da base de dados de treinamento da CNN divididas por colunas de classes (0 a 9, A, B, C, F, G e ruídos recorrentes) 44
Figura 16 – Exemplos de imagens em que o sistema não foi capaz de extrair a ID adequadamente. 48

LISTA DE TABELAS

Tabela 1 – Arquitetura usada na rede neural convolucional 43
Tabela 2 – Média e desvio padrão dos resultados obtidos por divisão com o sistema proposto. 46
Tabela 3 – Desempenho de cada filtro nas identificações bem sucedidas da avaliação do sistema proposto 47

LISTA DE ABREVIATURAS E SIGLAS

Auto ID	Identificação Automática
CNN	*Convolutional Neural Networks*
DL	*Deep Learning*
ID	Identificação
IA	Inteligência Artificial
IMS	*Intelligent Manufacturing Systems*
IoT	*Internet of Things*
LTQ	Laminador de Tiras à Quente
ML	*Machine Learning*
OCR	*Optical Character Recognition*
ReLU	*Rectified Linear Unit*
RFID	Radio Frequency Identification
RNA	Redes Neurais Artificiais
ROI	*Region of Interest*
UFES	Universidade Federal do Espírito Santo

LISTA DE SÍMBOLOS

f	imagem original
x	coordenada sobre o eixo horizontal da imagem
y	coordenada sobre o eixo vertical da imagem
b	elemento estruturante definido
T	imagem após filtragem *top-hat*
i	iluminância da imagem
r	refletância da imagem
z	logaritmo da imagem
s	imagem após filtragem homomórfica
d	exponencial da imagem após filtragem homomórfica
i_0	iluminância da imagem após filtragem homomórfica
r_0	refletância da imagem após filtragem homomórfica
H	filtro passa-altas para filtragem homomórfica
γ_H	parâmetro superior do filtro homomórfico
γ_L	parâmetro inferior do filtro homomórfico
p	número de pixels com uma determinada intensidade numa imagem
l	intensidades de cinza de uma imagem
t	limiar de Otsu
$\omega_{0,1}$	pesos para cálculo do limiar de Otsu
σ	variância das classes de intensidade da imagem
U, V	conjuntos não vazios
Q	correspondência entre cada elemento de U e V
g	mapa de características

j	classe num problema de classificação
K	número de classes num problema de classificação
$\mathbf{w_j}$	pesos do neurônio j
b_j	*bias* do neurônio j
x_r	ponto de referência para transformação geométrica
x_c	ponto de referência do caractere segmentado
θ_{ref}	ângulo de alinhamento do caractere

SUMÁRIO

1 INTRODUÇÃO . . . 14
1.1 A Manufatura Inteligente . . . 14
1.1.1 Sistemas *Auto ID* . . . 15
1.1.2 A Manufatura Inteligente no Contexto da Produção Siderúrgica . . . 16
1.2 Reconhecimento Óptico de Caracteres . . . 17
1.3 Justificativa . . . 18
1.4 Trabalhos Relacionados . . . 18
1.5 Escopo do Trabalho . . . 19
1.6 Objetivos . . . 21
1.6.1 Objetivo Geral . . . 21
1.6.2 Objetivos Específicos . . . 21
1.7 Estrutura do Texto . . . 21

2 REFERENCIAL TEÓRICO . . . 23
2.1 Reconhecimento Óptico de Caracteres . . . 23
2.2 Seleção da Região de Interesse . . . 24
2.3 Filtragem e Binarização . . . 24
2.3.1 Filtro *Top Hat* . . . 24
2.3.2 Filtro Homomórfico . . . 25
2.3.3 Algoritmo de Otsu . . . 27
2.4 Segmentação . . . 27
2.5 Transformações Geométricas . . . 28
2.6 Classificação das Imagens Segmentadas . . . 29
2.6.1 Redes Neurais Convolucionais . . . 29
2.6.2 Tesseract . . . 32
2.7 Pós-Processamento . . . 32

3 SISTEMA PROPOSTO . . . 33
3.1 Extração da Região de Interesse . . . 33
3.2 Filtragem e Binarização . . . 35
3.2.1 Filtro *Top Hat* . . . 35
3.2.2 Filtro Homomórfico . . . 35
3.2.3 Binarização . . . 37
3.2.4 Arquitetura de Filtragem . . . 37
3.3 Transformação Geométrica . . . 38
3.4 Remoção de Pequenos Componentes Conectados . . . 39

3.5 Alinhamento e Segmentação 40
3.6 Classificação 41
3.7 Pós-Processamento 42

4 RESULTADOS 44
4.1 Bancos de Imagens 44
4.2 Métricas de Desempenho 45
4.3 Resultados do Sistema 45

5 CONCLUSÕES E PROJETOS FUTUROS 49
5.1 Conclusões 49
5.2 Trabalhos futuros 50

REFERÊNCIAS 51

1 INTRODUÇÃO

1.1 A Manufatura Inteligente

A indústria manufatureira moderna está enfrentando vários desafios, caracterizados por mudanças rápidas, produtos tecnicamente mais complicados e pressões para aumentar a produção e eficiência ao mesmo tempo em que se reduz o desperdício de material. Exige-se mais qualidade dos produtos finais e, da mesma forma, a gestão do armazenamento e a da cadeia de suprimentos têm se tornado operações mais refinadas, exigindo atenção, desenvolvimento e melhorias contínuas especiais.

Nesse contexto surge a manufatura inteligente, que tem por objetivo otimizar a produção e as transações de produtos, fazendo pleno uso de informações avançadas e tecnologias de fabricação (ZHONG et al., 2017). É considerado como um novo modelo de fabricação baseado em ciência e tecnologia e soluções inteligentes que aprimora bastante o projeto, produção, gerenciamento e integração de todo o ciclo de vida de um produto típico.

Tais ciclos produtivos podem ser facilitados e acompanhados usando sensores inteligentes, modelos adaptativos de tomada de decisão, materiais avançados, dispositivos inteligentes e análise de dados (LI et al., 2017). A eficiência da produção, a qualidade do produto e o nível de serviço são, dessa forma, melhorados, aumentando a competitividade da indústria e aprimorando sua capacidade de enfrentar a dinâmica e as flutuações do mercado global (DAVIS et al., 2012).

Outro aspecto importante na manufatura moderna é a capacidade de rastrear e localizar objetos em tempo real ou quase real. O rastreamento de produtos e seu status através dos diferentes estágios de fabricação podem acelerar significativamente a produção e aumentar a eficiência e visibilidade da produção. Além disso, permite a melhoria nas etapas de planejamento no fluxo produtivo e na gestão de recursos empresariais (HODGSON; NABHANI; ZAREI, 2010).

Uma forma de realização deste modelo de fabricação é o uso de sistemas de manufatura inteligente (IMS, do inglês *"intelligent manufacturing systems"*), considerados a próxima geração de sistemas de manufatura, obtidos pela adoção de novos modelos, formas e metodologias para transformar os tradicionais sistemas de fabricação em sistemas inteligentes. A partir deles, todos dados sobre os processos e fluxos produtivos são obtidos e

disponibilizados para sistemas de produção considerados holísticos (WANG et al., 2006).

Dessa forma, a fabricação inteligente requer certas tecnologias adjacentes, a fim de sensoriar e atuar sobre a produção. Essas tecnologias alimentam os sistemas de controle de processos com informações em tempo real, permitindo assim que, dado o cenário, decisões sejam tomadas em tempo hábil (WAN et al., 2017). Algumas tecnologias também possuem inteligência artificial (IA), que permite que os sistemas de manufatura aprendam com as experiências para, então, realizar uma prática industrial inteligente e onipresente. Conceitos semelhantes a fabricação inteligente incluem fabricação e computação em nuvem, além de comunicação de objetos por meio da Internet das Coisas (*Internet of Things - IoT*) (LU; CECIL, 2016).

No contexto da Indústria 4.0, considerada a quarta revolução indústrial e caracterizada pela integração da manufatura a partir de Sistemas *cyber*-físicos, IoT e Internet dos Serviços, os IMS são a base para implantar tecnologias avançadas para criar processos e serviços de maior valor agregado.

1.1.1 Sistemas *Auto ID*

Nesse novo contexto produtivo integrado, o uso de tecnologias de apoio como sistemas de identificação automática (*Auto ID*) na fabricação pode ajudar a capturar o status em tempo real das manufaturas e fornecer informações importantes de produção para tomar decisões de alto nível de maneira rápida e flexível (ZHANG et al., 2015).

Essa tecnologia se refere a identificadores exclusivos e legíveis por máquina. O *Auto ID* refere-se aos métodos de identificação automática de objetos (ou seja, itens e pessoas), coleta de dados relacionados a objetos e entrada dos dados automaticamente em sistemas de computador a serem usados para uma finalidade específica. As tecnologias de identificação automática mais comuns são: códigos de barras, identificação por radiofrequência (RFID, do inglês *"radio frequency identification"*), biometria e reconhecimento de voz, faixas magnéticas, reconhecimento óptico de caracteres e tecnologia de *chip* inteligente.

As partes principais de um sistema típico de identificação automática são as seguintes (ZHAO; CAO; TU, 2014):

***Tag*:** Uma *tag* (ou etiqueta, do inglês) exclusiva de identificação é anexada a um objeto para distinguí-lo de objetos semelhantes. A *tag* pode ser um rótulo, microcircuito ou similar que contenha um código exclusivo e/ou outras informações. O objeto pode

ser um item como um componente, produto, ferramenta, dispositivo de fabricação ou uma pessoa, como um trabalhador.

Interrogador: Um leitor de *tags* que solicita o código de ID e outros possíveis dados armazenados e os interpreta automaticamente.

Servidor e banco de dados: O interrogador é conectado a um sistema de gerenciamento de informações que é usado para processar e armazenar as informações obtidas da *tag*. Isso geralmente inclui um servidor conectado a um banco de dados e conexão com sistemas de segundo plano.

Usuários finais: Os usuários finais têm acesso ao banco de dados por meio, por exemplo, de interface gráfica de usuário ou de sistemas de *Business Intelligence*, que permite a utilização de informações para diversos fins.

As técnicas de identificação automática possibilitam a implementação de monitoramento individual de materiais, peças e produtos, por exemplo, em diferentes estágios de fabricação. Além disso, a rastreabilidade das informações relacionadas à produção é importante nos estágios posteriores da cadeia de suprimentos, pois pode melhorar a qualidade do ciclo de vida do produto e até mesmo reduzir seu impacto ambiental, tornando as informações relacionadas ao material visíveis.

1.1.2 A Manufatura Inteligente no Contexto da Produção Siderúrgica

Nos últimos 20 anos, as indústrias siderúrgicas começaram a investir em soluções inteligentes e a maioria das empresas de grande porte aplicaram sistemas de gerenciamento e controle de informações (WANG et al., 2006). No entanto, esses sistemas dependem da disponibilidade de dados fundamentais de seus processos, adquiridos em tempo real, de forma que codificar e identificar as informações contidas nos materiais torna-se crucial para obter informações e realizar o gerenciamento da automação (BREWER; SLOAN; LANDERS, 1999).

Um requisito básico e de alta prioridade na indústria siderúrgica, por exemplo, é que o produto entregue ao cliente final tenha as mesmas características do produto pedido, considerando as diferentes ligas possíveis, durezas, tamanhos, pesos, valores e suas aplicações. Procura-se com isso garantir a segurança dada a aplicação do produto, reduzir custos devido a trocas de mercadorias e garantir a qualidade e competitividade da companhia. Para tanto, a maioria dos produtores adota soluções de identificação, como códigos de barras, identificação por radiofrequência, identificação impressa associada ao reconhecimento

óptico de caracteres (OCR, do inglês *"optical character recognition"*) ou tecnologias de *chips* inteligentes (LIUKKONEN; TSAI, 2016).

Inserido neste contexto, este trabalho foi desenvolvido na ArcelorMittal Tubarão, uma fábrica de produtos planos de aço: placas e bobinas. Nesta indústria, a laminação de bobinas ocorre em grande escala e de forma contínua, com produção aproximada de cerca de 500 bobinas por dia, ou cerca de uma bobina a cada 3 minutos. A grande produção exige uma logística refinada e rastreamento confiável, de forma que, nesta unidade, a identificação dos produtos é realizada por meio da impressão de sete caracteres - duas letras indicando o ano e o mês e cinco números - na frente das bobinas de aço. A identificação está associada a informações como a placa de origem, composição da liga, dimensões e instruções de armazenamento.

1.2 Reconhecimento Óptico de Caracteres

O reconhecimento óptico de caracteres (OCR) tem sido uma das principais tecnologias na aplicação do reconhecimento automático de padrões nos últimos 60 anos, e ainda é amplamente usado para reconhecer caracteres impressos em altas taxas. Aplicações comuns de OCR incluem rastreamento de data ou lote de mercadorias embaladas, diferentes aplicações de manipulação de documentos, e leitura de números de série de produtos. A técnica é tipicamente usada para digitalização de documentos ou livros e classificação automática de mensagens, por exemplo, mas também pode ser usada para identificação de objetos (LIUKKONEN; TSAI, 2016).

O OCR é baseado na simbologia 2D e o processo de reconhecimento inclui a varredura do texto, caractere por caractere, analisando cada imagem de caractere e traduzindo-as em um formato legível pela máquina (MORI; NISHIDA; YAMADA, 1999). A vantagem da técnica é que tanto máquinas quanto humanos são capazes de ler o mesmo texto. Por outro lado, o estágio de varredura requer um contato relativamente próximo com o objeto, a taxa de varredura é relativamente baixa e, em geral, a decodificação produz mais erros do que a varredura de código de barras.

1.3 Justificativa

A ocorrência de identificações erradas nas bobinas, ou a falta de identificações, pode acarretar: *a*) na entrega de produtos errados aos clientes, reduzindo a confiabilidade que os mesmos depositam na indústria; *b*) em multas e sanções devido ao não cumprimento de acordos/contratos; *c*) em custos adicionais por ter que remanufaturar o produto pela falta de informação de suas especificações; *d*) em custos adicionais em testes laboratoriais para identificar bobinas não marcadas; *e*) atraso nas linhas de acabamento ou no despacho, com possível custo adicional de utilização do Terminal Portuário Siderúrgico.

Portanto, é desejado um sistema que avalie e informe aos operadores sobre a qualidade das impressões realizadas na linha de laminação, e se elas são coerentes com as identificações esperadas.

Já existe uma estrutura física disponível para obtenção das imagens de interesse, incluindo a câmera fixa e a infraestrutura de comunicação entre a câmera e os sistemas de controle e supervisão, de forma que o uso do OCR neste contexto mostra-se uma solução barata e facilmente implantável.

1.4 Trabalhos Relacionados

Diferentes trabalhos propuseram o uso de OCR para verificar a qualidade de impressão em processos e ambientes industriais.

Zhao et al. (2014) discutem a marcação e o reconhecimento óptico de caracteres em placas de aço, o que requer um pré-processamento similar considerando o ambiente industrial e o material. O trabalho propõe diferentes filtros de câmera para diferentes temperaturas de placas, mas não trata de alinhamento de texto e aplica uma solução de binarização que não consegue lidar com imagens de baixo contraste. Para a classificação, utilizam a rede neural de propagação reversa (*backpropagation*), pois é vantajosa no processamento de dados não-lineares, tolerância a falhas e aprendizado autônomo, sendo adequada para o reconhecimento de caracteres impressos em aço.

Valveny e López (2003) estudam um sistema de reconhecimento numérico para controle de qualidade de etiquetas de sachês cirúrgicos, em um cenário com iluminação controlada, apesar das diferentes posições de impressão possíveis. Seu estágio de reconhecimento

apresenta uma solução robusta para lidar com numerais sobrepostos e quebrados usando características dos caracteres, mas não apresenta resultados com taxas de erro reais. A classificação é realizada extraindo medidas de características e então comparando com um dicionário de modelos, já que os caracteres sofrem pouca influência de ruído.

Patki et al. (2015) concentram-se na impressão matricial para classificar automaticamente caixas de papelão usando o filtro *Bottom Hat* (GONZALEZ; WOODS et al., 2002) como técnica de pré-processamento, que poderia tratar a variação de raios e uma extração robusta de recursos. Na classificação utiliza-se uma máquina de vetores de suporte (*support vector machine* - SVM). Os pesquisadores alcançaram uma precisão maior que 98%, mas não conseguiram tratar textos distorcidos.

A maioria das técnicas sobre textos curvos que necessitam de alinhamento é desenvolvida para imagens de documentos, utilizando as linhas paralelas como referência de posicionamento, como Oliveira et al. (2010) que desenvolveu uma técnica de fusão de linhas complexa e eficiente com taxas de precisão elevadas. Shejwal e Bharkad (2017) propuseram uma técnica mais simples baseada em caixas delimitadoras, que são organizadas usando as linhas de documentos como referência e tratadas para compensar os textos distorcidos, mas alcançaram uma precisão de 77.24%, coerente com o dificuldade da tarefa.

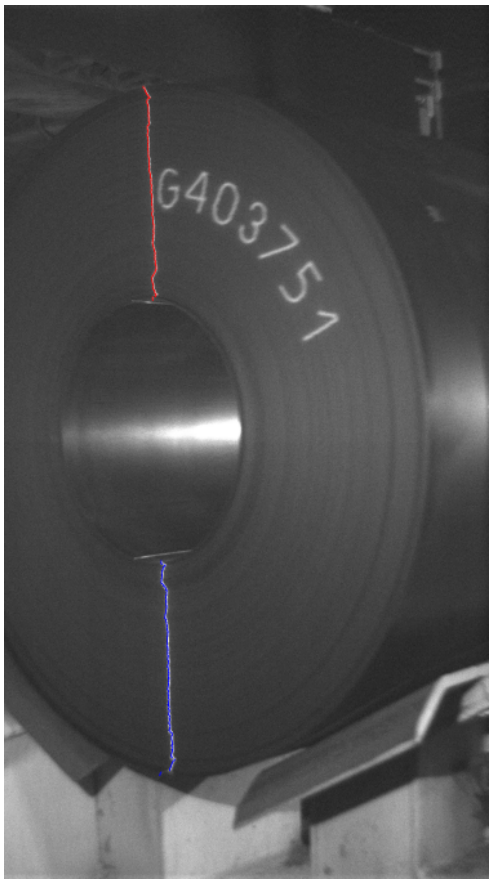
1.5 Escopo do Trabalho

O presente trabalho propõe um sistema de Reconhecimento Óptico de Caracteres (OCR) para extrair a identificação (ID) impressa em bobinas de aço, permitindo que se compare a identificação impressa àquela registrada nos sistemas de rastreamento da empresa, possibilitando intervenções no caso de impressões incorretas. Destaca-se, portanto, que o sistema tem acesso ao sistema de rastreamento, de forma que a identificação a ser encontrada na imagem é previamente conhecida, e deseja-se checar se de fato está presente e com qualidade.

Os possíveis erros que deseja-se detectar são a impressão da identificação incorreta (completa mas referente à bobina errada, ou incompleta), a ausência do ID e problemas na máquina de impressão. Um exemplo de identificação impressa é mostrado na Figura 1.

Como pode-se observar, a identificação é impressa em forma de arco, seguindo a curvatura da bobina, o que requer um processo de alinhamento antes da classificação. O ambiente industrial também requer robustez do sistema proposto em relação a grandes variações de

Figura 1 – Exemplo de bobina de aço com identificação impressa.



Fonte: ArcelorMittal Tubarão.

iluminação, ruído e especialmente características de impressão.

A solução proposta baseou-se nas etapas usuais dos sistemas de OCR: aquisição, pré-processamento, binarização, segmentação e classificação (MORI; NISHIDA; YAMADA, 1999). No entanto, diante das especificidades encontradas, algumas adaptações foram feitas. Foi observado que um único passo de pré-processamento de imagens não cobria todos os possíveis problemas a serem tratados. Portanto, neste trabalho foi proposta uma estratégia que utiliza diferentes tipos de filtros, de modo que, se um filtro não funcionar, o processo recomeça com outro filtro ou parâmetros. Este procedimento é válido neste trabalho porque há uma "resposta correta" para comparar com as descobertas do OCR.

1.6 Objetivos

1.6.1 Objetivo Geral

O objetivo geral deste trabalho é o desenvolvimento de um sistema que permita a extração por imagem da identificação impressa em bobinas de aço utilizando processamento digital de imagens (PDI) e aprendizado de máquina. Busca-se, dessa forma, garantir a rastreabilidade dos produtos produzidos no Laminador de Tiras à Quente (LTQ), na ArcelorMittal Tubarão, a fim de impedir que saiam dali bobinas com identificação errada ou inexistente. O sistema deve ser robusto o suficiente para lidar com variações significativas de iluminação e posicionamento das bobinas nas imagens.

1.6.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Realização de revisão bibliográfica referente a técnicas de processamento de imagens, aprendizado de máquina e trabalhos relacionados ao reconhecimento óptico de caracteres em ambiente industrial;
- Construção do banco de imagens a ser usado para testes e treinamento;
- Desenvolvimento e avaliação de desempenho do sistema proposto e documentação.

1.7 Estrutura do Texto

O presente trabalho está estruturado da seguinte maneira:

Capítulo 1: este capítulo inicial tem como objetivo contextualizar o trabalho e o problema aqui estudado, apresentando o problema e métodos sobre a solução do mesmo;

Capítulo 2: aqui são apresentados os fundamentos teóricos necessários para a compreensão do restante do trabalho;

Capítulo 3: neste capítulo é especificada a implementação do modelo escolhido e detalhado o seu funcionamento interno;

Capítulo 4: neste capítulo são apresentados o banco de dados e o *framework* de programação utilizados, bem como os resultados obtidos e uma breve análise dos resultados;

Capítulo 5: no capítulo final são apresentadas as conclusões e propostas de trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este capítulo tem como objetivo trazer os fundamentos teóricos necessários para a compreensão do restante do trabalho. As seções a seguir foram divididas em função das etapas necessárias para a realização do reconhecimento dos caracteres impressos nas bobinas de aço. A etapa de aquisição das imagens não será abordada já que ela não é realizada no presente trabalho.

2.1 Reconhecimento Óptico de Caracteres

O reconhecimento óptico de caracteres é a conversão de imagens de texto digitado, manuscrito ou impresso em texto codificado e interpretável por uma máquina.

É usado, por exemplo, na digitalização de documentos, na leitura de elementos textuais em fotos, como textos em fachadas ou placas de carros, ou mesmo para leitura de texto de legendas sobrepostos em uma imagem. É comumente usado como forma de entrada de dados a partir de registros em papel impresso - como documentos de passaporte, faturas, extratos bancários, recibos informatizados, entre outros (MORI; NISHIDA; YAMADA, 1999).

Em geral, seu processamento é dividido em algumas etapas principais, adaptadas de acordo com a aplicação desejada:

Aquisição: É o processo de captura da imagem a ser processada, sendo em geral realizada com algum tipo de câmera ou sensor óptico.

Extração da Região de Interesse: Nesta etapa delimita-se, dentro da imagem a ser processada, qual a região contém as informações úteis a serem lidas a partir de características como a localização, cor ou dimensões dos elementos da imagem.

Pré-Processamento: Envolve, em geral, o realce das regiões de interesse com intuito de tornar a informação mais facilmente processada. Costuma-se utilizar diferentes técnicas de filtragem e correção de contraste e brilho, além da binarização da imagem.

Segmentação: Neste estágio a informação a ser lida é separada do resto da imagem, o que pode ser feito, por exemplo, palavra a palavra ou caractere a caractere, dada a aplicação.

Classificação: É quando a informação da imagem será identificada, ou classificada de acordo com os possíveis caracteres correspondentes.

2.2 Seleção da Região de Interesse

A Região de Interesse (ROI, do inglês *"Region of Interest"*) é a região definida automaticamente a partir de parâmetros obtidos na própria imagem (ou por um usuário) onde o processamento estará totalmente concentrado (ALBUQUERQUE; ALBUQUERQUE, 2000).

No caso de problemas relacionados ao OCR, diferentes abordagens podem ser utilizadas para delimitar a ROI, por exemplo, pode-se definir a área de interesse como aquela imediatamente em torno dos caracteres, ou mesmo como o quadrante da bobina em que os caracteres são impressos.

Escolhendo-se uma ou outra abordagem, pretende-se obter uma região mais restrita cujo processamento deve ocorrer com menor custo computacional e com maior chance de sucesso na identificação dos caracteres.

2.3 Filtragem e Binarização

2.3.1 Filtro *Top Hat*

Aplicado, em geral, em imagens binárias (preto ou branco), o processamento morfológico modifica as áreas conectadas dentro da imagem (grupos de bit 0 ou 1) modificando seus contornos de acordo com a forma do elemento estruturante. A operação de erosão, por exemplo, permite separar objetos que se tocam. Ao contrário, a operação de dilatação permite preencher furos no interior de um objeto ou mesmo ligá-los.

Como as duas operações são iterativas é possível realizar uma sequência de N operações de erosão e dilatação sucessivas ou mesmo alternadas. A operação de erosão seguida de uma dilatação pelo mesmo elemento estruturante é conhecida como *abertura*, já a operação inversa é chamada de *fechamento*.

O *Top Hat* é um filtro morfológico usado principalmente para remover a iluminação desigual do fundo da imagem e destacar elementos e detalhes de ordem semelhante e de intensidade superior ao elemento estruturante definido (SOLOMON; BRECKON, 2011).

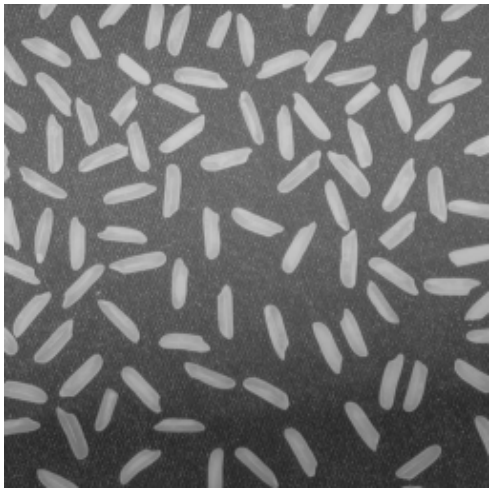
Define-se a operação realizada por este filtro como a diferença entre a imagem original e sua abertura morfológica assim como na Equação 2.1, em que $\circ$ define a operação de abertura, f é a imagem original e b é o elemento estruturante definido.

$$T(f) = f - f \circ b \tag{2.1}$$

As Figuras 2(a) e 2(b) mostram o tipo de resultado obtido com o uso do filtro.

Figura 2 – Exemplo de uso do filtro morfológico *Top Hat*

(a) Imagem original



(b) Imagem após uso do filtro

Fonte: Gonzalez, Woods et al. (2002)

2.3.2 Filtro Homomórfico

O filtro homomórfico é baseado no modelo de iluminância - refletância para corrigir a iluminação não uniforme em uma imagem. Este modelo se traduz como: a) quanta luz se concentra na cena observada (iluminância i) e b) quanta luz é refletida na cena (refletância r), como observado na Eq. 2.2.

$$f(x, y) = i(x, y)r(x, y) \tag{2.2}$$

Tenta-se separar os dois componentes, considerando que o primeiro geralmente tem variação espacial lenta, enquanto o último tende a variar abruptamente (GONZALEZ; WOODS et al., 2002). Para tanto, define-se $z(x, y)$ na Eq. 2.3 e encontra-se sua transformada de Fourier. Pode-se então aplicar $H(u, v)$, um filtro passa-altas mostrado na Figura 3, de forma que $S(u, v) = H(u, v)Z(u, v)$ e então trazê-lo para o domínio espacial, obtendo a Eq. 2.4. Como $z(x, y)$ foi obtido a partir do logaritmo da imagem original. A operação inversa produz o resultado esperado $d(x, y)$, onde $i_0 e r_0$ são respectivamente os componentes de iluminância e refletância da imagem, agora separáveis:

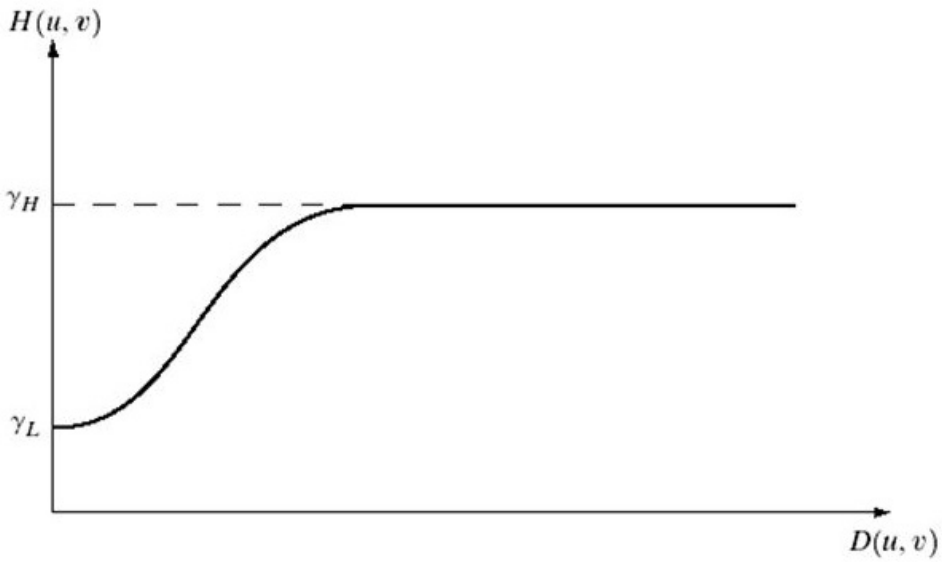
$$z(x, y) = \ln f(x, y) = \ln i(x, y) + \ln r(x, y) \tag{2.3}$$

$$s(x, y) = i'(x, y) + r'(x, y) \tag{2.4}$$

$$d(x, y) = \exp s(x, y) = i_0(x, y) + r_0(x, y) \tag{2.5}$$

Dessa forma, para reduzir a influência da iluminação, o Filtro Homomórfico configura-se como um filtro passa-altas no domínio da frequência, e portanto, as componentes de baixa frequência – associadas à iluminação – têm suas intensidades reduzidas. O filtro tem dois parâmetros que definem a forma de sua curva, que se comporta como mostrado na Figura 3: γ_L, que atenua as baixas frequências, e γ_H, que aumenta a intensidade das altas frequências, onde $D(u, v)$ está relacionado com a frequência e $H(u, v)$ é a intensidade.

Figura 3 – Seção transversal de um filtro homomórfico circular simétrico no domínio da frequência



Fonte: Gonzalez, Woods et al. (2002)

2.3.3 Algoritmo de Otsu

O algoritmo de Otsu baseia-se no pressuposto de que a imagem é uma combinação de duas classes de *pixels* seguindo um histograma bi-modal: primeiro plano (*foreground*) e o fundo (*background*).

Ele define o limiar ótimo como aquele que minimiza a variação intra-classe (OTSU, 1979). Para tanto, ele verifica todos os possíveis valores de limiar para a binarização procurando o valor em que a entropia seja a mínima. O método de Otsu determina o valor limite baseado na informação estatística da imagem onde, para um valor de limiar escolhido t, a variância dos agrupamentos do primeiro e segundo plano pode ser calculada.

O valor limite ideal é calculado minimizando a soma ponderada das variâncias $\sigma^2_{0,1}$ do grupo, em que os pesos ω_0 e ω_1 são calculados a partir das probabilidade do histograma $p(l)$ para cada valor de intensidade i dos respectivos grupos definida na Eq. 2.6. Os índices 0 e 1 definem os dois grupos divididos pelo limiar (OTSU, 1979).

$$\sigma^2_w(t) = \omega_0(t)\sigma^2_0(t) + \omega_1(t)\sigma^2_1(t) \tag{2.6}$$

Os pesos para uma imagem de 8 bits (256 níveis de intensidade) são calculados nas Equações 2.7 e 2.8. O limiar ótimo t é encontrado após avaliar todos os 256 níveis de intensidade.

$$\omega_0(t) = \sum_{l=0}^{t-1} p(l) \tag{2.7}$$

$$\omega_1(t) = \sum_{l=t}^{255} p(l) \tag{2.8}$$

2.4 Segmentação

Segmentação se refere ao processo de dividir uma imagem digital em múltiplas regiões (conjunto de *pixels*) ou objetos, com o objetivo de simplificar e/ou mudar a representação de uma imagem para facilitar a sua análise (GONZALEZ; WOODS et al., 2002). Dessa forma, pode-se partir para uma análise específica daquilo que foi selecionado.

Este processo se concentra, em geral, em usar determinadas características comuns a determinados grupos de *pixels* que o separem do resto da imagem. Para tanto, em geral utilizam-se técnicas para acentuar tais diferenças, como determinados filtros.

Em geral, a questão da segmentação é dividida em duas abordagens (SOLOMON; BRECKON, 2011; GONZALEZ; WOODS et al., 2002):

Métodos de contorno/fronteira Detecção de bordas como meio de identificação de fronteiras entre regiões, buscando, portanto, grandes diferenças entre *pixels*.

Métodos baseados em regiões Alocação de *pixels* a uma dada região com base no grau de similaridade entre eles (como no processo já visto de Binarização).

2.5 Transformações Geométricas

Uma transformação geométrica, ou espacial, é usada, por exemplo, no intuito de alinhar conjuntos de imagens que foram tiradas em momentos diferentes ou com diferentes sensores, corrigir distorções ou efeitos da orientação da câmera ou mesmo para adicionar efeitos especiais.

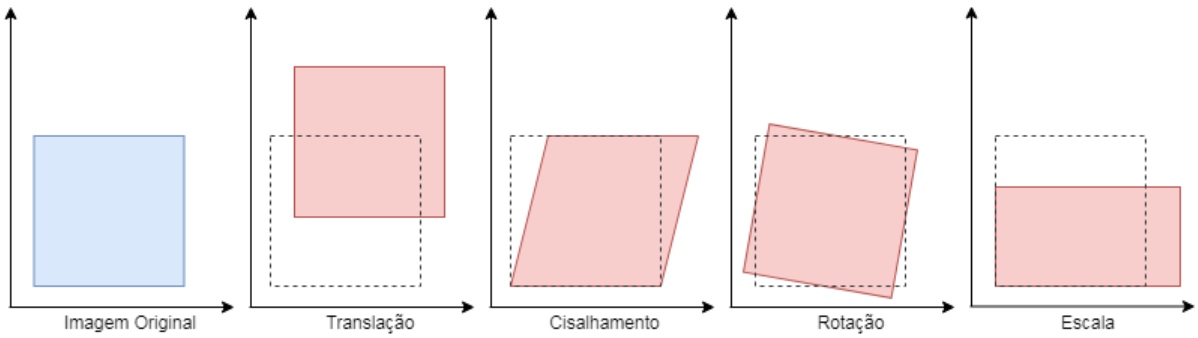
Os tipos de transformações mais comuns são chamadas de Transformações Afins, que mantêm o paralelismo das retas e modificam o sistema de coordenadas de imagem a partir de uma matriz de transformação. Exemplos de transformações afins são mudanças de escala, translação, rotação e cisalhamento (GONZALEZ; WOODS et al., 2002), como mostrado na Figura 4.

Formalmente, uma transformação afim de um conjunto não vazio U em um conjunto não vazio V é uma correspondência Q que, a cada elemento x de U, associa um único elemento $y = Q(x)$ de V e denota-se:

$$Q : U \to V \tag{2.9}$$

No contexto do OCR, as transformações geométricas são aplicadas principalmente no alinhamento dos caracteres antes que sejam inseridos em um classificador. No caso de leitura de caracteres impressos é comum que, em geral devido ao posicionamento da câmera e ao formato do material de fundo, os textos estejam deformados (MORI; NISHIDA; YAMADA, 1999). A abordagem mais comum envolve a segmentação dos conjuntos de

Figura 4 – Tipos de transformações afins



Fonte: Produção da própria autora.

caracteres e seu alinhamento subsequente usando a linha de base dos conjuntos segmentados como referência, aplicando uma transformação afim em cada conjunto separadamente.

2.6 Classificação das Imagens Segmentadas

Após o processo de segmentação, em que a informação desejada já se encontra separada do plano de fundo da imagem, passa-se ao processo de identificar os objetos obtidos. Existem diversas estratégias em torno da classificação automática, que em geral podem ser agrupadas em supervisionadas, em que se treina o sistema com exemplos que possam ser generalizados (ex. Redes Neurais Artificiais) e não supervisionadas, em que se utilizam apenas os dados disponíveis, em busca de características que dividam as classes de objetos de acordo com o objetivo do PDI (SOLOMON; BRECKON, 2011).

Existem alguns *softwares* que realizam o reconhecimento e conversão de caracteres em imagens, assim como o Tesseract, da Google. Inc., e que podem ser utilizados nessa etapa. Estes, no entanto, exigem adequações específicas da imagem no processo de pré-processamento, como alinhamento dos caracteres e binarização.

2.6.1 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNN) são um tipo de rede neural artificial (RNA) muito aplicada em tarefas relacionadas à Visão Computacional. Apesar de possuir um funcionamento semelhante às perceptron multicamadas – um tipo

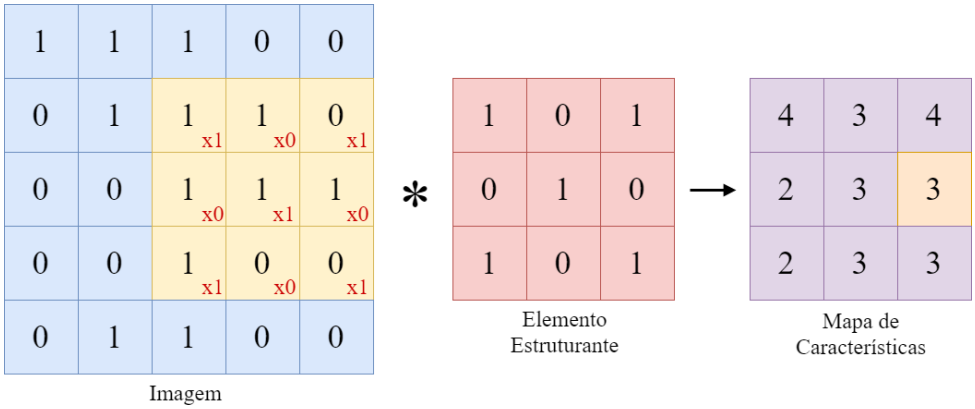
de RNA –, elas tem a capacidade de avaliar informações espaciais, recebendo por entrada uma estrutura – parte da imagem – e não um vetor. Os diferentes tipos de camadas da rede são operadores que tem por função capturar informações espaciais das estruturas de entrada:

Convolucional: É a camada que caracteriza este tipo de rede neural. Ela realiza uma convolução entre uma imagem $f(x, y)$ e um elemento estruturante $z(x, y)$ de tamanho $2l \times 2m$ definido pelo usuário, obtendo um mapa de características $g(x, y)$, como mostrado na Eq. 2.10.

$$g(x, y) = b(x, y) * f(x, y) = \sum_{s=-l}^{l} \sum_{t=-m}^{m} b(s, t) f(x - s, y - t) \tag{2.10}$$

A Figura 5 ilustra a tarefa de convolução realizada sobre a imagem representada em azul. O elemento estruturante representado em vermelho, também chamado de *kernel*, desliza sobre a imagem multiplicando seus elementos (mostrados em vermelho, no canto inferior direito das células da imagem) e resultando no elemento em laranja do mapa de características.

Figura 5 – Ilustração da operação de convolução 2D



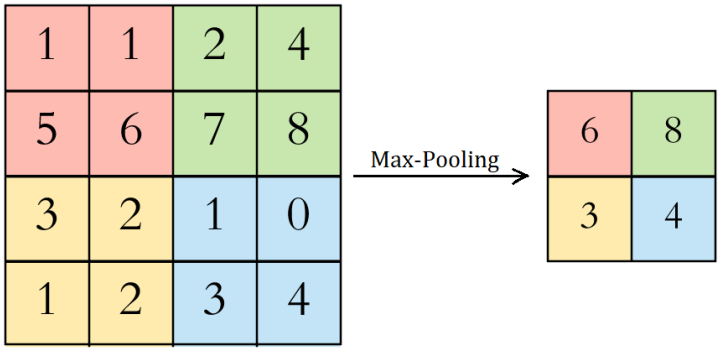
Fonte: Adaptado de Dertat (2017)

Ao longo do treinamento, apenas os pesos do elemento estruturante são atualizados. Assim a rede apresenta um aprendizado mais eficiente já que considera as vizinhanças, e melhor desempenho computacionalmente, já que apenas atualizam-se os pesos do elemento. Na prática, numa rede CNN profunda, entre centenas e milhares de elementos estruturantes são utilizados, distribuídos entre as várias camadas da rede.

***Pooling*:** Este operador tem por função principal reduzir as dimensões da entrada, aumentando a eficiência do treinamento ao descartar informações menos relevantes (BOUREAU; PONCE; LECUN, 2010). A operação de *pooling* também ajuda a tornar os mapas de características aproximadamente invariantes a pequenas translações da imagem de entrada (GOODFELLOW; BENGIO; COURVILLE, 2016).

De forma semelhante à convolução, no *pooling* se desliza um elemento estruturante sobre a matriz de entrada e seleciona-se um valor em cada sobreposição. No *max-pooling*, como mostrado na Figura 6, seleciona-se o maior elemento. No entanto, alternativamente, pode-se utilizar a média ou mediana.

Figura 6 – Ilustração da operação de *max-pooling*



Fonte: Adaptado de Dertat (2017)

***Dropout*:** Esse operador é introduzido na rede para evitar o *overfitting*. Para tanto, ele omite parte dos neurônios em cada amostra de treinamento zerando seus valores de ativação temporariamente (DOORN, 2014). Isso possibilita uma maior capacidade de generalização da rede, ao treinar os neurônios com combinações de entradas aleatórias.

***Flatten* e *Dense*:** Ao final de uma CNN é comum que se insira os dados finais – vetorizados pela camada de *flatten* –, em uma RNA completamente conectada, também chamada de *dense neural network* ou RNA densa.

Cada uma das camadas de convolução e da RNA utilizam funções de ativação no intuito de inserirem não linearidades à arquitetura:

ReLU: Do inglês *Rectified Linear Unit* (Unidade Linear Retificada), é aplicado ao mapa de características obtido da operação precedente de forma a retificar valores negativos como mostrado na Eq. 2.11. Na equação, $g(x, y)$ é o mapa de características de entrada, (x, y) são as coordenadas no mapa de características e a função $max[.]$ retorna o maior valor avaliado.

$$\mathrm{ReLU}(x, y) = \max\left[0, g(x, y)\right] \tag{2.11}$$

Softmax: Essa função atribui probabilidades de uma saída pertencer a cada classe em um problema multiclasse. Essas probabilidades decimais devem totalizar 1.0, o que auxilia o treinamento a convergir mais rapidamente (KIVINEN; WARMUTH, 1998).

A operação realizada é descrita matematicamente pela Eq. 2.12, que apresenta a probabilidade de uma entrada $\mathbf{x}$ na função *Softmax* ser associada a classe j, onde K é o número de classes e $\mathbf{w}_{j,k}$ e $b_{j,k}$ são, respectivamente, os pesos e *bias* do neurônio.

$$Softmax(y = j|\mathbf{x}) = \frac{e^{(\mathbf{w}_j^T \mathbf{x} + b_j)}}{\sum_{k \in K} e^{(\mathbf{w}_k^T \mathbf{x} + b_k)}} \tag{2.12}$$

2.6.2 Tesseract

O Tesseract é um mecanismo de reconhecimento de caracteres ópticos de código aberto, com capacidade de reconhecer mais de 100 idiomas, mantido pela Google Inc. Treinado principalmente para processar documentos impressos, ele tem melhores resultados em imagens binárias com regiões de texto poligonais definidas (SMITH, 2007).

O primeiro passo do processamento é uma análise de componentes conectados na qual os contornos dos componentes são armazenados (SMITH, 2007). Ao inspecionar o alinhamento de contornos e a hierarquia (quais estão contidos uns nos outros), detectam-se as regiões de textos. Os grupos de textos são organizados em linhas de texto, que são por sua vez divididas em palavras de acordo com o tipo de espaçamento de caracteres.

Em seguida, é feita uma primeira tentativa de reconhecer cada palavra por vez. Cada palavra que é satisfatória é passada para um classificador adaptativo como dados de treinamento, de forma que ao fim do documento o classificador adaptativo tem mais chances de reconhecer com mais precisão o texto. Para aproveitar-se do classificador, agora treinado, realiza-se uma segunda passagem pelo texto (SMITH, 2007). Uma fase final verifica se existem textos de outros tamanhos no documento.

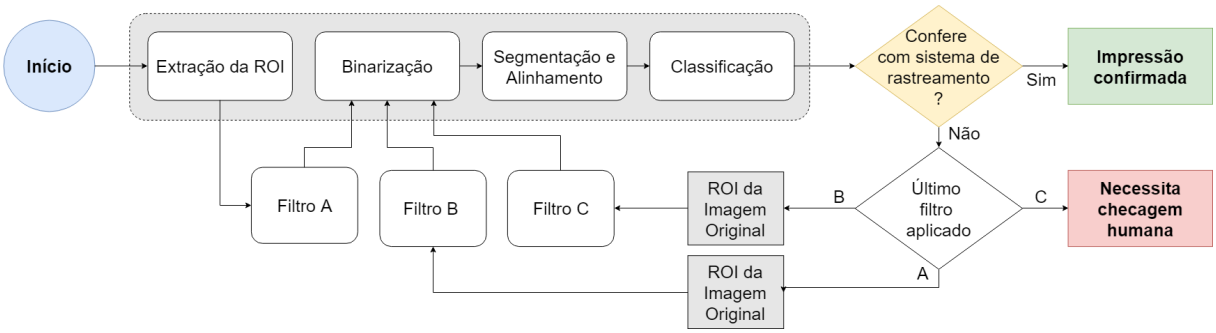
2.7 Pós-Processamento

Após realizadas todas as etapas de processamento, no contexto da OCR, o Pós-Processamento consiste em utilizar algum tipo de dado externo para conferir se a informação obtida da imagem pode estar correta. Um exemplo clássico é o uso de OCR para reconhecimento de placas veiculares, o que obrigatoriamente, quando trata-se de veículos no Brasil como modelo usado até 2018, deve obter uma informação constituída de 3 letras e 4 números. Pode-se, portanto, checar se o dado obtido está incorreto se o mesmo apresentar, por exemplo, o número errado de letras ou números ou se a ordem colocada está errada.

3 SISTEMA PROPOSTO

O sistema proposto baseia-se na arquitetura descrita pelo fluxograma mostrado na Figura 7. As seguintes seções do presente capítulo explorarão as ferramentas usadas por cada uma das etapas do processamento.

Figura 7 – Fluxograma do Sistema Proposto



Fonte: Produção da própria autora.

3.1 Extração da Região de Interesse

O primeiro passo do processamento é a extração da Região de Interesse (ROI). Este estágio é importante para diminuir a interferência de áreas claras e escuras da imagem nos caracteres desejados (como o reflexo central e a base da bobina e o fundo) no uso de técnicas de processamento que considerem toda imagem. Como existem vários tamanhos de bobinas, uma seleção de ROI adaptável foi desenvolvida com base no tamanho da bobina e ajustada empiricamente para que fosse o mais próximo possível do grupo de caracteres.

As imagens de entrada foram obtidas de um sistema de medição que utiliza um laser como referência para medir o formato da bobina. O laser projeta uma linha vertical no centro da bobina, que é digitalmente coberta com um marcador vermelho na parte superior e um marcador azul na parte inferior, como pode ser visto nas Figuras 8(a) e 8(b). Portanto, para obter uma referência de tamanho para determinar as dimensões do ROI, usamos informações desses marcadores.

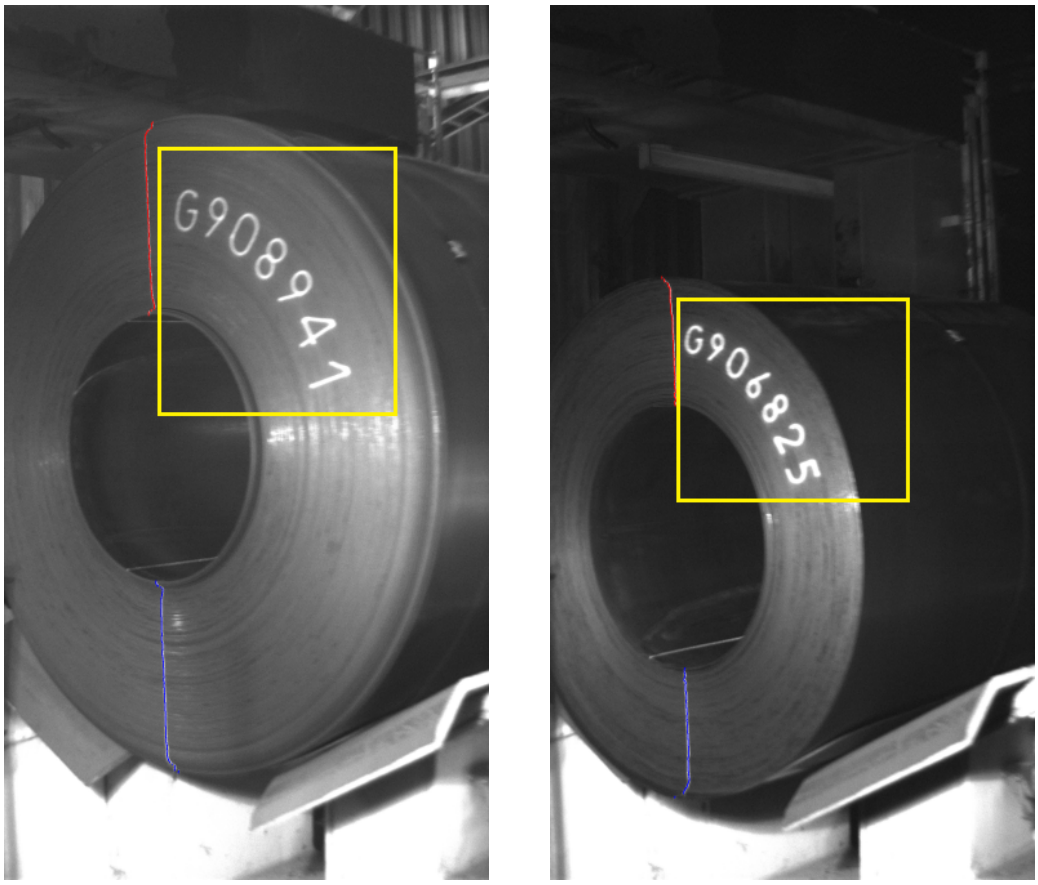
Para tanto, o marcador vermelho é selecionado pela sua cor, morfologicamente dilatado para conectar possíveis *pixels* desconectados, e seus pontos extremos superior e inferior são

obtidos. A largura da ROI é selecionada do ponto inferior até 200 *pixels* à sua direita, e da altura de 20 *pixels* abaixo do ponto superior até 70 *pixels* abaixo do ponto inferior. Os resultados são mostrados nas Figuras 8(a) e 8(b). Em seguida, as ROI são dimensionadas para que todas tenham a mesma altura (300 *pixels*), mantendo sua proporção.

No processo de selação, a largura é mantida a mesma pois, como a câmera é fixa, a distorção da imagem da bobina devido à aquisição oblíqua é mantido constante. Isso é usado em etapas subsequentes no ajuste geométrico da imagem para remoção da distorção.

Figura 8 – Seleção da região de interesse.

(a) ROI em bobina grande (b) ROI em bobina pequena



Fonte: Adaptado de ArcelorMittal Tubarão

A marcação em vermelho pode estar fragmentada devido a problemas no feixe de laser ou mesmo ao formato da bobina. Nesse caso, de todos os grupos de marcação encontrados, utiliza-se a marcação com os pontos mais próximos ao centro da bobina selecionando seu ponto mais inferior encontrado. Foi observado que estes pontos aparecem mais consistentemente, já que a deformação em torno do núcleo da bobina é geralmente menor. Assim, define-se a ROI sem a informação do tamanho da bobina, usando um valor padrão médio.

3.2 Filtragem e Binarização

As imagens das bobinas apresentam uma grande variabilidade em relação à iluminação, devido principalmente à interferência da luz solar e à desregulação ou ausência da iluminação artificial do local. Dessa forma, faz-se necessário uma filtragem específica para tratar a variação de iluminação, reduzindo sua influência. Para tratar este problema, o sistema está equipado com três filtros diferentes, de forma que os filtros subsequentes só são usados se os resultados obtidos pelo anterior não forem adequados, como mostrado na Figura 7. As próximas seções explorarão os filtros usados e por fim a arquitetura escolhida será discutida.

3.2.1 Filtro *Top Hat*

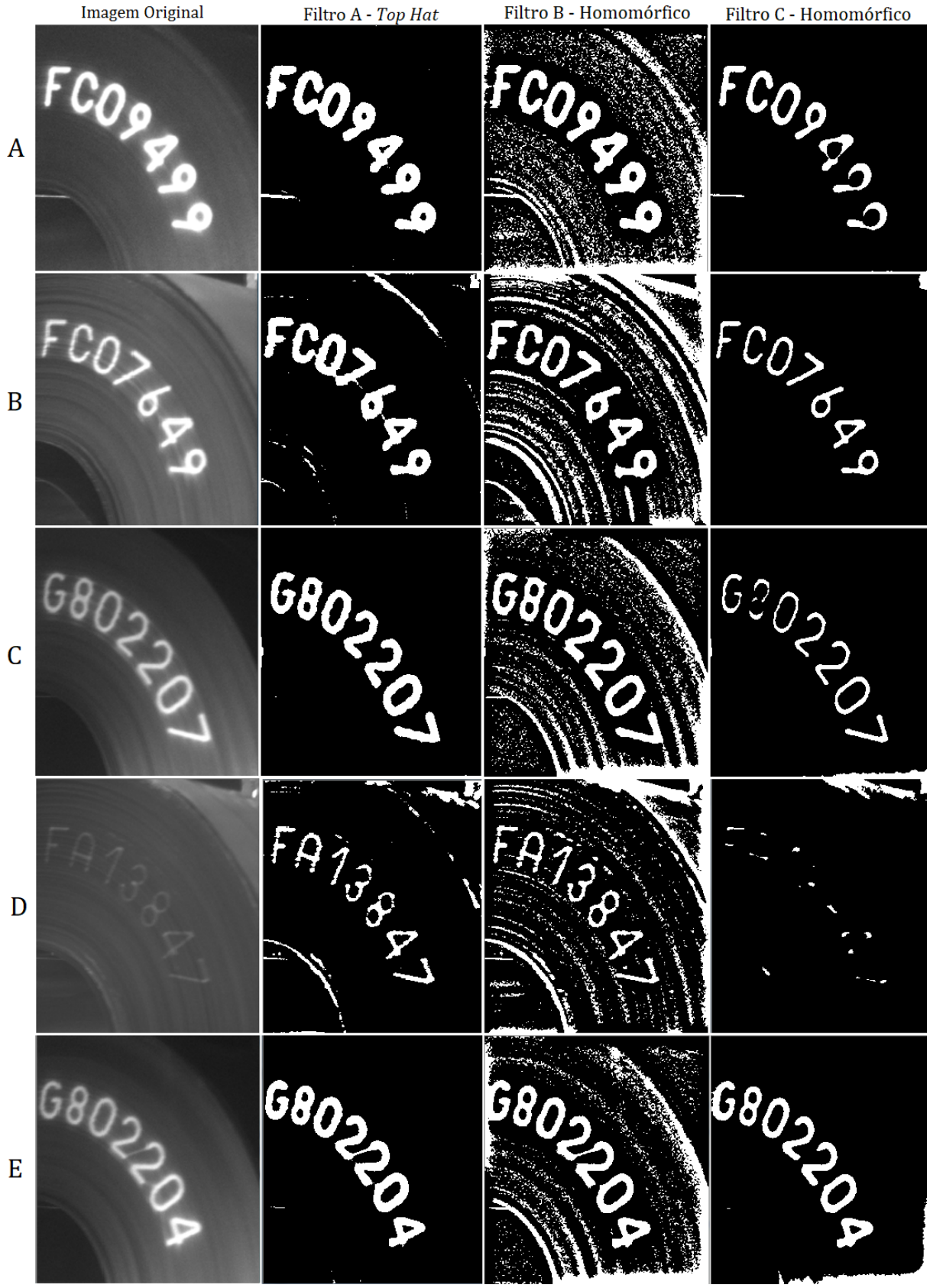
O primeiro filtro (Filtro A na Figura 7) é um *Top Hat* morfológico. Geralmente é utilizado para tratar o problema de iluminação não uniforme nas imagens, o que a torna útil para tratar os reflexos da bobina e do gradiente de luz. Nesta aplicação foi usado um elemento estruturante elíptico com eixo menor de valor 11 *pixels* e eixo maior 13 de *pixels* – tamanho este obtido empiricamente de forma que fosse semelhante ao dos caracteres, o que evita a perda de informação útil.

A segunda coluna da Figura 9 mostra os resultados de sua aplicação nas imagens da primeira coluna. Normalmente, as imagens de entrada possuem caracteres bem definidos, com contraste suficiente para que o filtro escolhido funcione bem.

3.2.2 Filtro Homomórfico

Existem casos em que o filtro *Top Hat* não obtém bom desempenho e, para estes casos, foi proposto o uso do filtro homomórfico com duas configurações diferentes. O primeiro, o Filtro B na Figura 7 ($\gamma_L = 0.05, \gamma_H = 2.5$), tem um γ_L muito baixo, permitindo uma alta sensibilidade a imagens com menor contraste, como o caso visto na quarta linha da Figura 9, mas inserindo uma grande quantidade de ruído. O último filtro, o Filtro C na Figura 7 ($\gamma_L = 0.5, \gamma_H = 2.8$), tem um γ_L dez vezes maior, permitindo lidar com um ruído mais brilhante, como é o caso da última linha da Figura 9, por exemplo, onde a máquina de impressão conectou os caracteres 2 e 2.

Figura 9 – Comparação do uso dos filtros após a binarização de Otsu mostrando em cada linha um diferente tipo de imagem encontrada. O filtro *Top Hat* filtrou adequadamente apenas a imagem da linha C, enquanto todos os outros exemplos necessitaram dos filtros B e C homomórficos.



Fonte: Produção da própria autora.

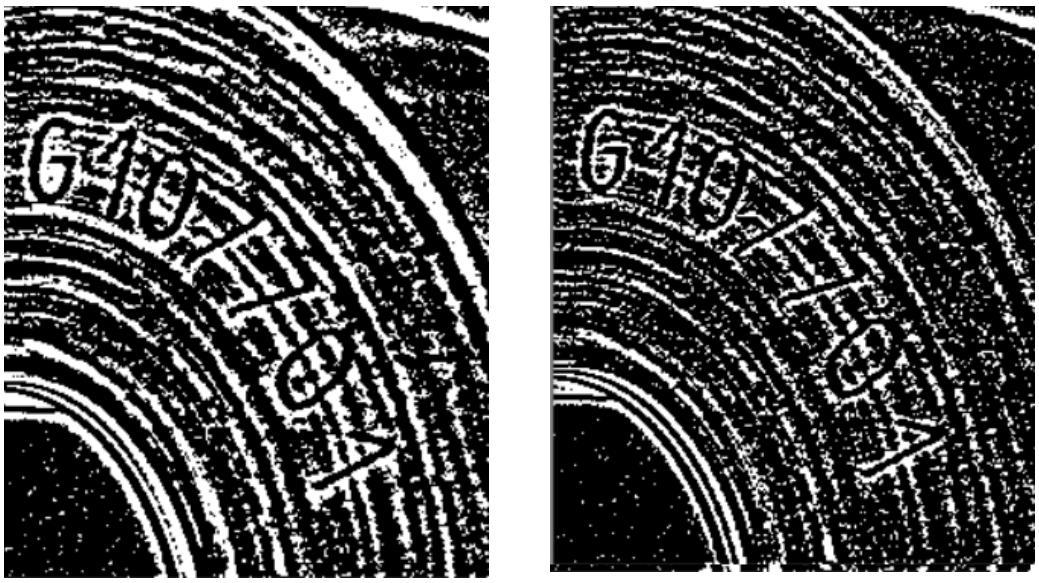
3.2.3 Binarização

Os filtros anteriores foram aplicados para aumentar o contraste global, a fim de obter uma imagem bimodal. Após a filtragem, independentemente de qual filtro foi usado, a próxima etapa é aplicar um limiar na imagem para obter uma imagem binária, para que os caracteres sejam legíveis e bem definidos após esse estágio. Para encontrar um limiar para cada imagem foi utilizado o método de Otsu.

O método de Otsu realiza uma binarização global, no entanto, existem métodos adaptativos que consideram a configuração local. Foram testadas a binarização adaptativa gaussiana, que calcula a média ponderada da vizinhança onde os pesos são referentes à uma janela gaussiana, e a binarização adaptativa de média, que se baseia na média simples da vizinhança. Os filtros, no entanto, inseriram uma grande quantidade de ruído, fundindo os caracteres, e principalmente sem aproveitar o contraste destes em relação ao *background*. As Figuras 10(a) e 10 exemplificam o efeito de tal tipo de binarização e sua grande inserção de ruído.

Figura 10 – Exemplos de utilização da binarização adaptativa.

(a) Adaptativa de média (b) Adaptativa gaussiana.



Fonte: Produção da própria autora.

3.2.4 Arquitetura de Filtragem

Na Figura 9 é mostrada a comparação dos resultados dos filtros escolhidos aplicados a imagens com características. Na primeira coluna da Figura 9 são exibidas as imagens

originais, na segunda coluna os resultados do filtro *Top Hat* (Filtro A) e nas duas últimas colunas os resultados dos Filtros Homomórficos (Filtros B e C), respectivamente. Como pode ser observado, cada caso exige um filtro diferente, de forma que e a aplicação de um filtro em determinada cena pode inclusive deteriorar a informação desejada, impedindo o uso de uma única estratégia de filtragem, mesmo com filtros adaptativos ou de vizinhança próxima.

O filtro *Top Hat* apresenta bom desempenho na maioria dos casos, no entanto existem determinados cenários em que outra estratégia se faz necessária. Em geral, quando a iluminação da cena mostra-se inadequada obtêm-se imagens muito escuras ou muito claras e brilhantes ou com a identificação impressa esmaecida. Nas imagens mostradas na quarta linha da Figura 9, pode-se observar como a utilização dos filtros homomórficos permitem cobrir, por exemplo como neste caso, imagens de contraste muito reduzido entre o fundo e os caracteres.

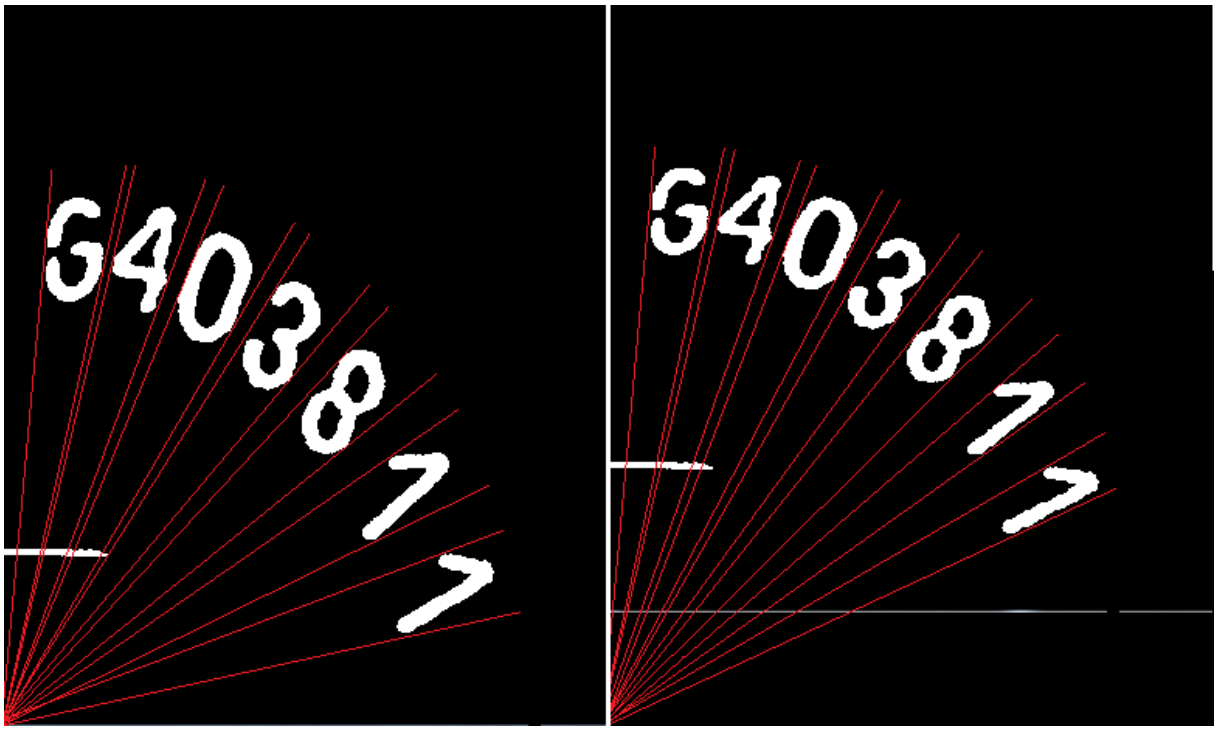
Como mostrado no fluxograma da Figura 7, adota-se então uma arquitetura de tentativa e erro, já que procura-se na imagem uma informação já conhecida. Um erro, portanto, significa que o sistema falhou em extrair a identificação conhecida ou que ela não está correta ou presente. Assim, um filtro subsequente só é usado se os resultados obtidos pelo anterior não forem adequados, configurando uma arquitetura sequencial que realiza todo o processamento sequencialmente com diferentes estratégias.

Outra opção seria um processamento paralelo, no qual todos os filtros seriam aplicados separadamente e o resultado mais bem-sucedido seria usado. Essa abordagem foi testada e, no entanto, descobriu-se que, como consome mais memória, passa a consumir também mais tempo. Além disso, observou-se que o filtro *Top Hat* apresentou baixo custo computacional e alta acurácia justificando, portanto, seu uso como filtro principal, o que diminui o tempo de processamento médio.

3.3 Transformação Geométrica

Embora as bobinas e os caracteres impressos tenham uma forma cilíndrica, de frente circular, a câmera fixa responsável pela aquisição da imagem não está alinhada com o centro da bobina, capturando uma visão elíptica do produto e deformando os caracteres. Felizmente, a câmera está no mesmo nível do centro, portanto, uma maneira simples de reverter a deformação é reduzir a imagem na dimensão y ou aumentá-la na dimensão x.

Figura 11 – Transformação geométrica para compensar deformação devido à aquisição oblíqua: À esquerda, imagem binarizada, à direita, imagem após a transformação e o preenchimento na base.



Fonte: Produção da própria autora.

Pode-se observar nas Figuras 8(a) e 8(b), que mostram a seleção da região de interesse de acordo com o tamanho da bobina, que quanto maior a bobina, maior é a deformação dos caracteres. Dessa forma, para fazer um ajuste proporcional às dimensões da bobina, se a altura original da ROI é maior que 200 *pixels*, a imagem é transformada em um quadrado, reduzindo-a verticalmente. Se a altura for menor que 200 *pixels* a transformação não é realizada, já que a deformação não é expressiva.

No caso em que a transformação é feita, adiciona-se também um preenchimento na base da ROI para manter o seu tamanho original, pois o canto esquerdo pode funcionar como ponto de referência de alinhamento. A Figura 11 mostra a imagem antes e depois da transformação, com linhas vermelhas adicionadas manualmente para mostrar como o procedimento é capaz de recuperar o arco circular original.

3.4 Remoção de Pequenos Componentes Conectados

Para finalizar o estágio de pré-processamento do sistema proposto, os componentes conectados restantes com conectividade 4 na imagem binária são avaliados por seu tamanho,

Figura 12 – Da esquerda para a direita: Imagem original; imagem filtrada pelo Filtro B (homomórfico) e binarizada; imagem redimensionada e sem pequenos componentes conectados.

Fonte: Produção da própria autora.

eliminando aqueles menores que o tamanho médio de um caractere. Como pode-se observar na Figura 12, o primeiro filtro homomórfico geralmente deixa uma quantidade considerável de ruído, que consegue ser tratado com eficácia neste estágio. É importante mencionar que possíveis caracteres segmentados podem ser filtrados por essa rotina, portanto, um dos filtros anteriores deve ser capaz de lidar com esse tipo de ocorrência.

3.5 Alinhamento e Segmentação

Após as etapas anteriores, o próximo procedimento é extrair os caracteres e alinhá-los, para que possam ser identificados. Para tanto, os componentes conectados em cada imagem binária são extraídos e sua ordem (da esquerda para a direita) é mantida para o reconhecimento de ID. Para remover componentes conectados que não são caracteres, eles são filtrados por diferentes características, tais como tamanho, proporção de altura $\times$ largura e região da imagem.

Usando o centro do círculo circunscrito aos componentes filtrados, um ponto de referência é obtido para cada um deles. Como discutido na Seção 3.3 sobre Transformação Geométrica, o canto inferior esquerdo pode ser usado como referência de alinhamento. Portanto, para alinhar os caracteres, o ângulo θ_{ref} entre o canto e cada centro do círculo é calculado usando a Equação 3.1. É importante lembrar que a origem do sistema de coordenadas se localiza no canto superior esquerdo, o que explica porque o canto inferior esquerdo de referência é $(x_r, y_r) = (0, h_{imagem})$, em que a h_{imagem} refere-se à altura total da imagem, enquanto o centro do caractere é $(x_c, y_c) = (x, y)$. Com esta informação, o ângulo é calculado usando

Eq. 3.2.

$$\theta_{ref} = arctan\left(\frac{y_r - y_c}{x_r - x_c}\right) - \frac{\pi}{2} \tag{3.1}$$

$$\theta_{ref} = arctan\left(\frac{h_{imagem} - y}{0 - x}\right) - \frac{\pi}{2} \tag{3.2}$$

Os pontos centrais dos caracteres, suas dimensões e os ângulos calculados são usados para extrair os retângulos alinhados com o ponto de referência (x_r, y_r), como pode ser visto na Figura 13. Depois disso, uma imagem negativa é criada concatenando os componentes cortados e alinhados, como mostrado na Figura 14. As imagens segmentadas são avaliadas para eliminar possíveis ruídos que não foram filtrados e passados como um caractere, e somente as imagens segmentadas de caracteres com mais de 1250 *pixels* pretos são inseridos no classificador, a serem explicados na seção a seguir.

Figura 13 – Retângulos inscritos alinhados à referência

Fonte: Produção da própria autora.

3.6 Classificação

Deve-se, em seguida, realizar a classificação das imagens segmentadas entre os possíveis caracteres. Foram avaliadas duas diferentes técnicas de classificação: o Teseract (SMITH,

Figura 14 – Imagem final dos caracteres alinhados

Fonte: Produção da própria autora.

2007), e uma rede neural convolucional (CNN).

O Tesseract é um mecanismo de reconhecimento óptico de caracteres de código aberto, com capacidade de reconhecer mais de 100 idiomas, e pode ser treinado para reconhecer outros (SMITH, 2007). No entanto, o sistema foi treinado principalmente para lidar com palavras, especialmente aquelas escritas em fontes específicas, e isso prejudica a identificação correta dos caracteres impressos.

Por sua vez, a CNN tem apresentado resultados adequados para diferentes tarefas relacionadas a imagens. Neste trabalho, as CNN foram treinadas para classificar as imagens entre as 17 classes possíveis: todos os números (10), as letras A, B, C, F, G, T (6), e uma classe de ruído, que reúne componentes segmentados que não são caracteres, mas possuem características semelhantes a eles e, portanto, não foram filtrados no estágio de segmentação,

A arquitetura proposta foi baseada em uma rede projetada para o *CIFAR-10 Task* (KUMAR, 2018), que envolve a classificação de 16 mil imagens do conjunto de dados CIFAR-10 (KRIZHEVSKY; NAIR; HINTON, 2018) em 10 categorias. Um resumo das camadas da rede neural convolucional, já adaptadas à classificação dos caracteres, pode ser encontrado na Tabela 1.

A CNN empregada neste trabalho usa o otimizador estocástico de gradiente descendente com uma taxa de aprendizado de 0,01 e 20 épocas. A CNN treinada é então usada para identificar os caracteres segmentados individualmente.

3.7 Pós-Processamento

O estágio de pós-processamento permite que o sistema compare a ID real com a saída do OCR e julgue se as informações correspondem. Neste estágio também podem ser tratados alguns possíveis erros de processamento, como ruídos detectados como caracteres. Na

Tabela 1 – Arquitetura usada na rede neural convolucional

Camada	**Configuração**
Convolutional 2D	Número/Dimensão dos *Kernels*: 16 (3, 3) Ativação: *Rectified Linear Unit*
Dropout	Valor de *Dropout*: 0.2
Convolutional 2D	Número/Dimensão dos *Kernels*: 16 (3, 3) Ativação: *Rectified Linear Unit*
Max Pooling 2D	Tamanho de *Pooling*: (2, 2)
Convolutional 2D	Número/Dimensão dos *Kernels Kernels*: 32 (3, 3) Ativação: *Rectified Linear Unit*
Max Pooling 2D	Tamanho de *Pooling*: (2, 2)
Flatten	
Dropout	Valor de *Dropout*: 0.2
Dense	Número de Neurônios: 512 Ativação: *Rectified Linear Unit*
Dropout	Valor de *Dropout*: 0.2
Dense	Número de Neurônios: 17 Ativação: *Softmax*

Fonte: Produção da própria autora.

aplicação proposta, é suficiente que os cinco caracteres finais se correspondam, já que os dois primeiros caracteres marcam o ano e o mês da instrução, respectivamente.

Para tanto, são definidas regras para lidar com identificações e ruídos incompletos no final ou no início da saída do OCR. Por exemplo, se há uma bobina na qual o ID no sistema de rastreamento é "FA12345", e o sistema proposto o lê como "FA12345", há uma correspondência. Por outro lado, se a saída for, por exemplo, "A12345", "12345", "FA123451", "A123451" ou "123451", o sistema também a aceitará como uma correspondência.

Se o sistema está operando com o primeiro ou segundo filtro, e não há uma correspondência, o sistema passa para o próximo filtro, reiniciando os procedimentos com a região original de interesse, como é mostrado na Figura 7. Se houver uma correspondência, a execução será interrompida e a imagem binária final dos caracteres alinhados será armazenada para que se possa verificar o resultado se necessário. Se após os três filtros não houver correspondência, a imagem original será armazenada com a imagem final resultante do último filtro, para que um operador humano possa avaliar manualmente as imagens. Essa separação entre os erros de impressão ou os erros do sistema OCR foi considerada para avaliar a precisão do sistema.

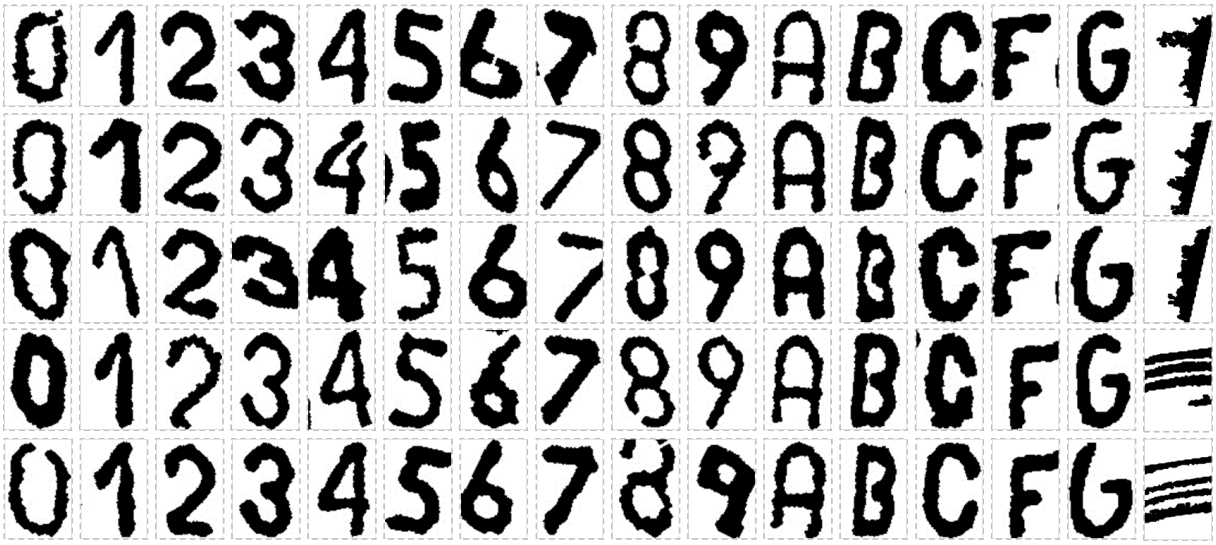
4 RESULTADOS

4.1 Bancos de Imagens

Para avaliar o sistema proposto, uma base de imagens cedida pela empresa foi utilizada. Este conjunto, organizado por data, reúne imagens de dimensões 376×679 de todas bobinas laminadas desde setembro de 2017 a outubro de 2018, com uma média de 500 bobinas por mês. Aproximadamente 3% de todas as imagens apresentam algum tipo de erro de impressão, contraste ou de iluminação que torna o ID ilegível.

Foram construídos, a partir deste banco, dois subconjuntos. O primeiro, um conjunto de imagens binarizadas dos caracteres, foi construído reunindo exemplos de cada mês que foram pré-processados de modo que os caracteres foram extraídos e então usados para treinar a CNN. Aproximadamente 1.500 imagens de cada caractere foram obtidos, totalizando cerca de 24.000 imagens de tamanho 180×120, como mostrado na Figura 15.

Figura 15 – Exemplos de imagens da base de dados de treinamento da CNN divididas por colunas de classes (0 a 9, A, B, C, F, G e ruídos recorrentes)



Fonte: Produção da própria autora.

O segundo, denominado conjunto de teste, foi utilizado para avaliar o sistema, sendo constituído por amostras das imagens originais. Este subconjunto é composto por cerca de 22.000 imagens distribuídas ao longo de todos os meses disponíveis, com $1701{,}3 \pm 164{,}9$ imagens por mês. Nenhuma dessas imagens foi usada no treinamento da CNN.

4.2 Métricas de Desempenho

Como mostrado no sistema da Figura 7, as imagens que correspondem à ID no Sistema de Rastreamento são rotuladas como "identificações bem sucedidas", enquanto as que não correspondem à ID são rotuladas como "identificações com falha" e devem passar por uma inspeção manual. As "identificações com falhas" podem ser causadas por problemas na impressão ou rastreamento, ou pela incapacidade do sistema de extrair e reconhecer a identificação. Diante disso, para avaliar o sistema proposto, definem-se dois possíveis tipos erros que constituem as falhas detectadas:

Erros de processo: Casos em que o sistema não consegue ler o ID porque ocorreu um erro de impressão, iluminação ou de rastreamento.

Erros do sistema de OCR: Casos em que o sistema não consegue ler a ID na imagem, mas um ser humano pode lê-lo.

O total de imagens testadas é a soma de "identificações bem sucedidas" e "identificações com falha". A soma dos "erros de processo" e dos "erros do sistema de OCR", por sua vez, constituem as identificações com falha.

Também foram divididos os erros do sistema de OCR: *a*) filtragem e segmentação e *b*) classificação, para que se pudesse avaliá-los separadamente. Os estágios de filtragem e segmentação foram mantidos juntos porque o último depende completamente do desempenho do primeiro. Uma métrica adicional foi utilizada:

Acurácia: Porcentagem de casos que o sistema de OCR proposto obteve sucesso, isto é, o número de acertos sobre a quantidade total de imagens do conjunto com ID legível. É portanto a proporção entre as identificações bem sucedidas e o total de imagens subtraídos dos "erros de processo".

4.3 Resultados do Sistema

O sistema proposto foi desenvolvido em Python, utilizando principalmente a biblioteca OpenCV2 e a biblioteca Keras para treinamento da CNN. Os teste descritos nesta seção foram executados em um computador com processador Intel i5 de 2,90GHz de processamento, 4GB de memória RAM e placa de vídeo Intel HD *graphics* 4600 *onboard*.

Inicialmente, avaliou-se o desempenho da CNN no reconhecimento dos caracteres. Foi realizada uma validação cruzada estratificada de cinco *folds* sobre o CNN usando o conjunto de dados de caracteres, obtendo uma acurácia média de $99,37 \pm 0,11\%$. Observa-se que, o alto valor de acurácia alcançado pela CNN indica que seu uso no sistema proposto é promissor.

A melhor CNN treinada obtida na validação cruzada foi utilizada no sistema. A fim de garantir a confiabilidade sobre a avaliação do sistema proposto, o conjunto de testes foi dividido em 10 partes estratificadas de acordo com os meses disponíveis. Assim, garante-se que cada parte contenha diferentes cenários de luz ocorridos ao longo do ano, além de todos os caracteres possíveis referentes aos meses. Portanto, o sistema foi executado 10 vezes para que, em cada execução, uma parte diferente fosse avaliada. Após 10 execuções, foram obtidas a média e o desvio padrão de cada medida de desempenho.

Na Tabela 2, os resultados experimentais são mostrados de acordo com os tipos de erros definidos na Seção 4.2. Foram testadas duas técnicas diferentes de classificação: CNN e Tesseract. Como visto, o primeiro alcançou melhor desempenho e alta taxa de precisão, o suficiente para ser integrado ao sistema de rastreamento. O tempo médio de processamento em ambas as soluções é suficiente para tratar o problema, porque cada bobina leva até três minutos para ser laminada. Também pode-se observar que, usando a CNN, os erros do sistema de OCR se devem principalmente aos estágios de filtragem e segmentação, que não foram capazes de lidar com certos problemas de contraste e iluminação. Por outro lado, usando o Tesseract, a etapa de classificação apresentou baixa acurácia e foi a principal causa dos erros do sistema.

Tabela 2 – Média e desvio padrão dos resultados obtidos por divisão com o sistema proposto.

Informação por divisão	**CNN** **Média ± Desvio Padrão**	**Tesseract** **Média ± Desvio Padrão**
Imagens	2211,3 ± 0,42	2211,3 ± 0,42
Identificações bem sucedidas	2091,8 ± 7,04	1793.9 ± 15,98
Erros de processo	77,4 ± 10,76	77,4 ± 10,76
Erros do sistema de OCR	42,1 ± 8,76	340 ± 19,5
Erros de filtragem e segmentação	91,25 ± 4,02%	9,42 ± 3,19%
Erros de classificação	8,75 ± 4,02%	88,94 ± 3,19%
Acurácia	98,02 ± 0,52%	84,62 ± 2,89%
Tempo por Imagem (s)	0,4151 ± 0,0168	0,6408 ± 0,0441

Fonte: Produção da própria autora.

Usando a CNN, que obteve melhor desempenho, foram avaliadas quais das diferentes camadas de filtro foi capaz de pré-processar a imagem para obter uma identificação bem

sucedida, com qualidade suficiente para que o OCR pudesse ser realizado. A Tabela 3 reúne os valores percentuais em que cada filtro obteve sucesso, de tal modo que foram desconsiderados os casos em que todos os filtro falharam. Porém, como observado na Tabela 2, o percentual de falha de todos os filtros é de aproximadamente 2%.

Tabela 3 – Desempenho de cada filtro nas identificações bem sucedidas da avaliação do sistema proposto

Filtro Usado	**Média**	**Desvio Padrão**
Filtro A - *Top Hat*	95,33%	0,65%
Filtro B - Homomórfico	3,11%	0,47%
Filtro C - Homomórfico	1,55%	0,36%
Total	100%	–

Fonte: Produção da própria autora.

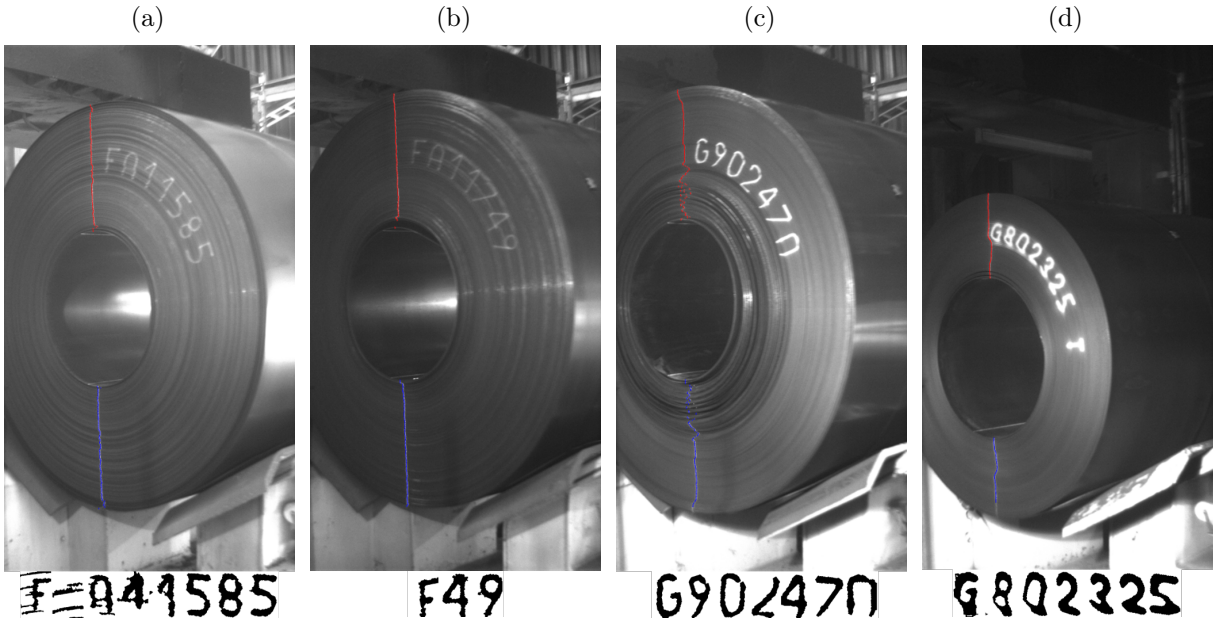
Conforme explicado no fluxograma da Figura 7, o segundo filtro é usado apenas quando o primeiro falha. Assim, como mostrado na Tabela 3, o Filtro *Top Hat* tem um alto índice de sucesso, sendo o único utilizado em 95,33% das imagens, portanto, este resultado justifica seu uso como filtro principal. Pode-se observar que o segundo e terceiro filtros são utilizados em apenas 4,66% dos casos, porém, este valor representa 1040 imagens considerando todo o conjunto de dados, um número expressivo de imagens que exigiriam uma análise do operador, demandando esforço humano e tempo para ser executado.

As imagens mostradas na Figura 16 exemplificam casos em que o sistema não foi capaz de realizar a extração da identificação, apesar de sua presença na imagem. Mostra-se também, abaixo de cada imagem, o resultado obtido após a filtragem, binarização, segmentação e alinhamento. Pode-se observar que, apesar de legíveis, a qualidade apresentada da impressão ou do cenário aumentam a complexidade da tarefa e exigem um tratamento ainda mais robusto do que o apresentado no presente trabalho.

Nas Figuras 16(a) e 16(b) são destacadas situações em que o contraste entre as letras e o material da bobina mesclam-se por apresentarem intensidades de cinza semelhantes, principalmente devido à iluminação sobre as espiras da bobina. O resultado é a grande inserção de ruído, que por ser identificado como caractere como na Fig. 16(a), ou que une os caracteres, que acabam sendo excluídos por apresentarem proporção ou dimensões erradas, como na Fig. 16(b). Outro problema de comum ocorrência é mostrado na Fig. 16(c), em que as espiras atrapalham a impressão da identificação, e a classificação falha. Já na Fig. 16(d), nota-se como a forma "telescópica" da bobina pode tornar a identificação deformada e de difícil processamento.

Apesar da ocorrência de falhas, os resultados experimentais comprovaram que o sistema proposto é capaz de tratar efetivamente o problema de identificação de caracteres, de

Figura 16 – Exemplos de imagens em que o sistema não foi capaz de extrair a ID adequadamente.



Fonte: ArcelorMittal Tubarão e a própria autora

modo que reduz o esforço humano para realizar essa tarefa.

Acrescenta-se que a arquitetura proposta permite a adição de outros filtros em sequência, se necessário, para lidar com outros cenários, com cautela quanto ao aumento do tempo de processamento. Destaca-se que a maioria dos erros do sistema ocorreu em imagens sequenciais devido aos problemas no braço robótico que imprime a identificação. Embora a identificação em alguns casos seja legível para um ser humano, a qualidade de impressão não foi ideal, portanto, este tipo de observação pode ser útil para indicar quando o sistema robótico precisa de manutenção.

5 CONCLUSÕES E PROJETOS FUTUROS

5.1 Conclusões

O objetivo principal deste trabalho foi o desenvolvimento de uma proposta de sistema de OCR para extrair, alinhar e reconhecer caracteres impressos em bobinas de aço em um ambiente industrial, a fim de verificar a qualidade do processo de impressão. A solução proposta deveria ser robusta o suficiente para lidar com variações significativas de iluminação e posicionamento das bobinas nas imagens.

A partir dos resultados obtidos e discutidos no Capítulo 4, constata-se a aplicabilidade do sistema proposto vinculado aos sistemas de rastreamento de produtos da empresa. Observou-se que ele é capaz de lidar com diversas variações na qualidade de iluminação e impressão, com um estágio robusto de pré-processamento. Os métodos de alinhamento e segmentação, por sua vez, são capazes de filtrar efetivamente os caracteres e alinhá-los, e a classificação baseada em uma rede neural convolucional pode reconhecer os caracteres, apesar de todo o ruído e distorções encontradas.

Além disso, a arquitetura do sistema proposto permite a adição de outros filtros em sequência, se necessário, para lidar com outros cenários, com cautela quanto ao aumento do tempo de processamento.

Entende-se que as CNN tem mostrado grande potencial para atividades de classificação, e que neste trabalho poderiam ter exploradas sem o uso de etapas de pré-processamento. No entanto, redes mais complexas como a que seria necessária para tal tarefa demandam grande custo computacional, que não era disponível, mas que talvez melhorasse os resultados obtidos.

Finalmente, conclui-se que foi obtido um protótipo de baixo custo e alta eficácia, que pode ser usado de forma a garantir a rastreabilidade. Observou-se também que a maioria dos erros do sistema ocorreu em imagens sequenciais, em que mesmo que a identificação fosse legível ao olho humano, a qualidade da impressão estava ruim, o que pode ser usado como indício para manutenção do braço robótico.

Um fator limitante é o fato de que, com a mudança do ano uma nova letra passa a ser

usada na marcação, que só poderia ser classificada com o retreinamento da CNN. Todavia, a etapa de pós-processamento foi pensada de forma que não se dependesse inteiramente do primeiro caractere, referente ao ano. Assim, propõe-se que a rede seja retreinada com as imagens obtidas a partir da virada do ano.

5.2 Trabalhos futuros

Um projeto futuro que teria grande impacto na eficácia do sistema seria a melhoria da qualidade na etapa de aquisição das imagens, melhorando a qualidade de iluminação e diminuindo a interferência de luz solar. Além disso, no que se refere ao escopo deste trabalho, possíveis melhorias são a adequação de outros filtros concatenados, para que possam abranger mais cenários, e a proposta de novas arquiteturas da CNN e hiperparâmetros.

Em desenvolvimentos futuros, o sistema proposto poderia ser usado, também, para detectar a necessidade de manutenção do braço robótico usado para imprimir os caracteres. Destaca-se que, dado o potencial observado em novos trabalhos usando redes neurais convolucionais, uma possível abordagem seria o treinamento da CNN para cada uma das tarefas intermediárias realizadas neste trabalho, como na segmentação dos caracteres. Outra possibilidade usando esta ferramenta seria o treinamento de uma CNN para a classificação tendo por entrada a segmentação dos caracteres sobre a imagem original, de forma que a rede seria responsável por lidar com os diferentes cenários.

REFERÊNCIAS

ALBUQUERQUE, M. P. de; ALBUQUERQUE, M. P. de. Processamento de imagens: métodos e análises. *Rio de Janeiro, Brasil*, v. 12, 2000. Citado na página 24.

BOUREAU, Y.-L.; PONCE, J.; LECUN, Y. A theoretical analysis of feature pooling in visual recognition. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. [S.l.: s.n.], 2010. p. 111–118. Citado na página 30.

BREWER, A.; SLOAN, N.; LANDERS, T. L. Intelligent tracking in manufacturing. *Journal of intelligent manufacturing*, Springer, v. 10, n. 3-4, p. 245–250, 1999. Citado na página 16.

DAVIS, J. et al. Smart manufacturing, manufacturing intelligence and demand-dynamic performance. *Computers & Chemical Engineering*, Elsevier, v. 47, p. 145–156, 2012. Citado na página 14.

DERTAT, A. *Applied Deep Learning - Part 4: Convolutional Neural Networks*. 2017. Disponível em: <https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>. Acesso em: 2018-12-07. Citado 2 vezes nas páginas 30 e 31.

DOORN, J. van. Analysis of deep convolutional neural network architectures. 2014. Citado na página 31.

GONZALEZ, R. C.; WOODS, R. E. et al. *Digital image processing*. [S.l.]: Prentice Hall New Jersey, 2002. Citado 5 vezes nas páginas 19, 25, 26, 27 e 28.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado na página 30.

HODGSON, S.; NABHANI, F.; ZAREI, S. Aidc feasibility within a manufacturing sme. *Assembly Automation*, Emerald Group Publishing Limited, v. 30, n. 2, p. 109–116, 2010. Citado na página 14.

KIVINEN, J.; WARMUTH, M. K. Relative loss bounds for multidimensional regression problems. In: *Advances in neural information processing systems*. [S.l.: s.n.], 1998. p. 287–293. Citado na página 31.

KRIZHEVSKY, A.; NAIR, V.; HINTON, G. *CIFAR-10 and CIFAR-100 datasets*. 2018. Disponível em: <https://www.cs.toronto.edu/~kriz/cifar.html>. Acesso em: 2018-09-26. Citado na página 42.

KUMAR, A. *Achieving 90% accuracy in Object Recognition Task on CIFAR-10 Dataset with Keras: Convolutional Neural Networks*. 2018. Disponível em: <https://appliedmachinelearning.blog/2018/03/24/>. Acesso em: 2018-09-26. Citado na página 42.

LI, B.-h. et al. Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering*, Springer, v. 18, n. 1, p. 86–96, 2017. Citado na página 14.

LIUKKONEN, M.; TSAI, T.-N. Toward decentralized intelligence in manufacturing: recent trends in automatic identification of things. *The International Journal of Advanced Manufacturing Technology*, Springer, v. 87, n. 9-12, p. 2509–2531, 2016. Citado na página 17.

LU, Y.; CECIL, J. An internet of things (iot)-based collaborative framework for advanced manufacturing. *The International Journal of Advanced Manufacturing Technology*, Springer, v. 84, n. 5-8, p. 1141–1152, 2016. Citado na página 15.

MORI, S.; NISHIDA, H.; YAMADA, H. *Optical character recognition*. [S.l.]: John Wiley & Sons, Inc., 1999. Citado 4 vezes nas páginas 17, 20, 23 e 28.

OLIVEIRA, D. M. et al. A new method for text-line segmentation for warped documents. In: SPRINGER. *International Conference Image Analysis and Recognition*. [S.l.], 2010. p. 398–408. Citado na página 19.

OTSU, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, IEEE, v. 9, n. 1, p. 62–66, 1979. Citado na página 27.

PATKI, S. N.; JOSHI, M.; KULKARNI, A. N. Dot matrix text recognition for industrial carton classification. In: IEEE. *Industrial Instrumentation and Control (ICIC), 2015 International Conference on*. [S.l.], 2015. p. 777–782. Citado na página 19.

SHEJWAL, M. A.; BHARKAD, S. D. Segmentation and extraction of text from curved text lines using image processing approach. In: IEEE. *Information, Communication, Instrumentation and Control (ICICIC), 2017 International Conference on*. [S.l.], 2017. p. 1–5. Citado na página 19.

SMITH, R. An overview of the tesseract ocr engine. In: IEEE. *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*. [S.l.], 2007. v. 2, p. 629–633. Citado 2 vezes nas páginas 32 e 42.

SOLOMON, C.; BRECKON, T. *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. [S.l.]: John Wiley & Sons, 2011. Citado 3 vezes nas páginas 25, 28 e 29.

VALVENY, E.; LÓPEZ, A. Numerical recognition for quality control of surgical sachets. In: IEEE. *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*. [S.l.], 2003. p. 379–383. Citado na página 18.

WAN, J. et al. A manufacturing big data solution for active preventive maintenance. *IEEE Transactions on Industrial Informatics*, IEEE, v. 13, n. 4, p. 2039–2047, 2017. Citado na página 15.

WANG, X. et al. Intelligent scheduling system of steelmaking and continuous casting based on erp/mes/pcs. In: IEEE. *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*. [S.l.], 2006. v. 2, p. 7381–7384. Citado 2 vezes nas páginas 15 e 16.

ZHANG, Y. et al. Real-time information capturing and integration framework of the internet of manufacturing things. *International Journal of Computer Integrated Manufacturing*, Taylor & Francis, v. 28, n. 8, p. 811–822, 2015. Citado na página 15.

ZHAO, Q.-J.; CAO, P.; TU, D.-W. Toward intelligent manufacturing: label characters marking and recognition method for steel products with machine vision. *Advances in Manufacturing*, Springer, v. 2, n. 1, p. 3–12, 2014. Citado 2 vezes nas páginas 15 e 18.

ZHONG, R. Y. et al. Intelligent manufacturing in the context of industry 4.0: a review. *Engineering*, Elsevier, v. 3, n. 5, p. 616–630, 2017. Citado na página 14.