

**DINCON 2017**CONFERÊNCIA BRASILEIRA DE DINÂMICA, CONTROLE E APLICAÇÕES
30 de outubro a 01 de novembro de 2017 – São José do Rio Preto/SP

Reconhecimento do ID das Placas Destinadas ao Enfornamento do Laminador de Tiras a Quente Utilizando Visão Computacional

Erick Ramos dos Santos¹

Graduando de Engenharia de Controle e Automação, IFES, Serra, ES,

Gustavo Maia de Almeida²

Departamento de Engenharia de Controle e Automação, IFES, Serra, ES

Resumo. Este artigo apresenta um método de reconhecimento, por meio de visão computacional, do ID das placas direcionadas aos fornos do LTQ (laminador de tiras a quente) de uma Siderurgia, sendo coletadas imagens destas placas através de uma câmera posicionada antes da entrada dos fornos. O atual sistema da empresa não consegue se manter eficaz durante todo o período do dia, devido à dificuldade de extrair características das imagens com iluminação não-uniforme. Com isso, um novo método é proposto, realizando um pré-tratamento com filtro de mediana e filtro gaussiano. As imagens tratadas são direcionadas a uma rede neural *Deep Learning* com arquitetura convolucional visando extrair as características e classificar os dígitos.

O novo código obteve resultados significativos, com uma taxa de acerto de 96,33% o que mostra o seu enorme potencial para aplicação.

Palavras-chave. Visão Computacional, *Deep Learning*, Rede Neural Convolucional.

1 Introdução

O controle do pátio de placas do laminador de tiras à quente de uma Siderurgia é feito através de um sistema semi-automatizado que, ao receber as informações específicas de cada placa, consegue movê-las de maneira inteligente e eficaz. O tipo de material, a espessura e o comprimento da placa, são exemplos destas informações. O recebimento destes dados faz-se necessário para realizar a movimentação das placas, pois baseado no pedido da bobina do cliente, o software decide qual placa deve ser enfornada, enviando assim, o comando ao operador.

Ao serem formadas no lingotamento contínuo, as placas recebem a escrita do seu respectivo ID em sua superfície, como representado na figura 1.

¹ rdserick@gmail.com

² gmaia@ifes.edu.br



Figura 1: ID escrito na placa

Ao sair do lingotamento contínuo, existem três destinos possíveis para a placa: pátio de placas do LTQ, condicionamento, em caso de placas com defeito que precisam ser recondiçionadas, e o pátio de placas destinadas para venda.

Quando a placa é direcionada à laminação de tiras a quente para a formação da bobina, uma câmera posicionada antes da entrada dos fornos fotografa a escrita na placa, visando confirmar se a placa que está sendo enforcada condiz com o ID informado pelo sistema. Essa averiguação do sistema é uma forma de controlar o processo e verificar sua confiabilidade, visto que se uma placa for enforcada de forma equivocada, a bobina produzida terá propriedades diferentes das exigidas pelo cliente, causando um atraso na produção da empresa e, consequentemente, um prejuízo notável.

Tendo em vista a forma de operação do sistema, o ID da placa é uma informação única e vital para o bom funcionamento de todo o processo da indústria. Logo, buscando uma maneira automática de realizar a confirmação do ID, um software de reconhecimento foi criado por meio de visão computacional. Desta forma, elimina a necessidade da presença de um operador durante todo o dia.

Este artigo está dividido em cinco seções. A seção 2 descreve em mais detalhes a problematização do processo de leitura dos ID's das placas. A seção 3 apresenta a metodologia adotada e os seus desafios. Os resultados obtidos são apresentados na seção 4, e por fim, a seção 5 apresenta as conclusões do estudo e propõe um caminho futuro de pesquisa.

2 Problematização

Atualmente, o sistema de reconhecimento não consegue se manter eficaz durante todo o período do dia, devido à dificuldade no controle da iluminação do amplo local onde estão situados os fornos. Outros problemas, como a falta de padrão de escrita nos lingotamentos contínuos e o vapor proveniente de uma placa com alta temperatura, também são válidos de serem citados, porém o último caso é desconsiderado, tendo em vista que identificar o ID nesta situação se torna algo impossível ao olho nu. Este sistema trabalha com pré-processamento da imagem tendo um algoritmo que a binariza através de *Adaptive Threshold* [11] [12], buscando sempre manter as bordas dos dígitos, visto que os números serão extraídos da imagem utilizando *Tesseract*, um *software* de OCR (Reconhecimento Óptico de Caracteres) de código aberto [10] que imprime os números de uma imagem em um arquivo txt. Ao ocorrer os problemas citados, realizar um pré-processamento via *software* em que a definição das bordas não é perdida, torna-se algo inviável a ser trabalhado, como ilustrado na figura 2, tornando o classificador ineficaz, possuindo uma taxa de acertos durante o dia em torno de 88%.



Figura 2: Imagem com iluminação ruidosa seguido da binarização feita com *Adaptive Threshold*

Portanto, torna-se necessário uma nova abordagem para a realização do reconhecimento do ID, de tal maneira que os obstáculos sejam vencidos via codificação.

Buscando vencer a não uniformidade do ambiente, o presente trabalho propõe um pré-tratamento das imagens, utilizando filtro de mediana e filtro gaussiano para diminuição de ruídos na binarização das mesmas. Juntamente com a melhoria no tratamento da imagem, é proposto um diferente classificador construído com *Deep Learning* tendo como arquitetura uma rede convolucional para extrair as características dos dígitos [9]. Desta forma, espera-se que o novo *software* seja mais versátil aos problemas, se mantendo eficaz durante todo o período do dia.

3 Metodologia

Como pode ser visto na figura 1, cada ID é composto por uma letra e nove dígitos. A letra representa a origem da placa, ou seja, qual dos três lingotamentos contínuos existentes na siderurgia produziu a placa, sendo esta informação irrelevante nesta parte do processo. Com isso, ao reconhecer a letra, o código irá ignorá-la, aumentando a eficiência do classificador, pois desta forma, evita a possibilidade de o classificador ‘confundir’ letras e dígitos semelhantes, como por exemplo, ‘B’ e ‘8’.

3.1 Codificação

As imagens captadas pela câmera estão em RGB, logo, o primeiro passo é a conversão da imagem em RGB para escala de cinza. Após a conversão, o próximo passo consiste na aplicação dos filtros para redução dos ruídos, tendo em vista que a aplicação da binarização sem um tratamento correto da imagem tende a ser algo ineficaz.

O filtro de mediana permite a atenuação dos ruídos da imagem através da ordenação dos níveis de cinza em ordem crescente de intensidade [2]. O ponto negativo deste filtro é a eliminação de linhas consideravelmente finas, o que não se torna algo prejudicial ao presente trabalho.

Buscando a suavização da imagem, a aplicação do filtro gaussiano torna-se uma alternativa, visto que se trata de um filtro passa-baixa, eliminando os valores de alta frequência como em [6]. Sendo uma suavização mais delicada, o filtro de média, consegue preservar mais os contornos da imagem, conseguindo assim, uma melhora na imagem sem perdas de informações importantes.



Figura 3: Binarização da imagem com aplicação dos filtros

Após realizado o pré-tratamento na imagem, e posteriormente sua binarização, como ilustrado na figura 3, torna-se possível o reconhecimento dos dígitos. Para separar cada dígito é utilizada a função do *Matlab* chamada *'bwlable'*, que retorna todos os componentes reconhecidos na imagem binarizada. Buscando o corte de cada dígito para que sejam submetidos ao classificador, utiliza-se a função *'regionprops'*, na qual obtém informações de cada componente. Recebendo a informação *'BoundingBox'* desta função, as seguintes características dos componentes são obtidas: as coordenadas do ponto inicial, largura e altura. Com essas informações, utilizando a função *'imcrop'*, a letra e os nove dígitos são cortados e separados em uma célula. Como já citado, a letra é irrelevante nesta parte do processo, logo, será eliminada, restando apenas os dígitos a serem reconhecidos, como mostra a figura 4.



Figura 4: Números recortados da imagem binarizada

3.2 Classificador *Deep Learning*

Todo sistema que envolve reconhecimento de padrões necessita de um classificador bem configurado, pois este que irá rotular cada informação de entrada. A rede neural artificial do tipo *Deep Learning* é uma técnica eficaz, com grande precisão, que atualmente, é considerado o estado da arte na área de visão computacional, possuindo ótimos resultados na classificação de imagens [3].

O *Deep Learning* é um método de *Machine Learning* que possui a capacidade de aprendizagem em grandes quantidades de dados de forma supervisionada, não supervisionada ou por meio de reforço [7]. Neste projeto, o método abordado é o supervisionado, visto que ao montar o banco de dados de treinamento da rede, será informado à rede o “rótulo” de cada dígito.

O banco de dados foi criado com dígitos retirados de algumas placas, para que a rede neural aprenda o padrão de escrita e consiga responder de maneira mais eficaz. Foram fornecidas cem imagens diferentes de cada dígito, ou seja, um banco de dados de mil exemplos. Quanto maior for o banco de dados de treinamento, mais robusta e eficaz a rede neural se torna. Cada dígito presente no banco possui as dimensões 30x50, ou seja, ao realizar o corte dos dígitos da placa, faz-se necessário o redimensionamento do dígito para as mesmas dimensões dos dados de treinamento, antes de ser classificado, conforme figura 5.

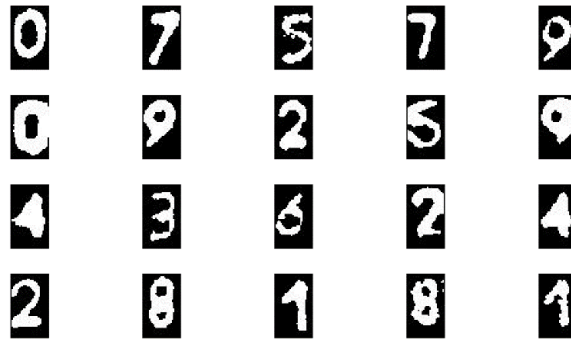


Figura 4: Exemplo de imagens presentes no banco de dados

Outra característica importante que deve ser definida ao trabalhar com *Deep Learning*, é a arquitetura da rede neural. Para cada tipo de dado, aconselha-se um tipo de arquitetura, e no caso do reconhecimento de padrões em imagens, a arquitetura que proporciona melhores resultados é a rede neural convolucional, possuindo trabalhos realizados no reconhecimento de caracteres manuscritos [4], na detecção de pedestres [8] e faces [5]. Este tipo de rede neural foi construído se baseando no córtex visual de animais, tendo em vista que ao receber um estímulo visual na retina, a imagem é reconhecida devido a regiões do cérebro que identificam características específicas do sinal passado [1].

Como é mostrado na figura 6, camadas de convolução são aplicadas na imagem de entrada. Estas camadas são compostas de neurônios que estão conectados a um conjunto de pixels da imagem, tendo cada conexão um peso diferente. Os neurônios aplicam filtros em pedaços específicos da imagem, produzindo uma saída que é passada para a próxima camada, tendo em vista que na CNN, o neurônio se conecta apenas a um subconjunto de neurônios da camada anterior, abordagem diferente das redes neurais clássicas. O conjunto de neurônios de uma mesma camada é chamado de mapa. Após as camadas de convolução, é comum utilizar uma camada de agregação (*pooling*) que tem como função a redução das dimensões da imagem (altura e largura), enquanto a profundidade é aumentada, aumentando assim, a velocidade de treinamento.

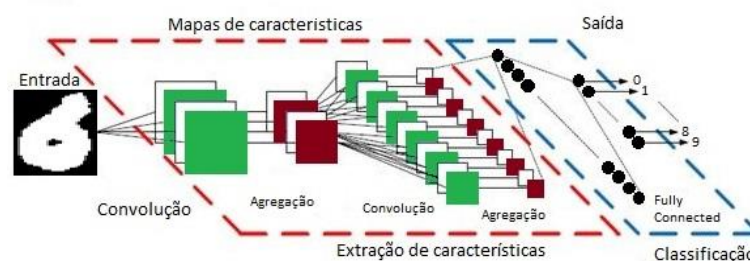


Figura 5: Arquitetura de uma rede convolucional

Ao final, é característico da arquitetura possuir uma camada de classificação visando rotular a resposta da rede se baseando nas saídas dos filtros.

Para criar a rede neural, algumas informações precisam ser determinadas, como a taxa de aprendizado (*learning rate*), número de neurônios, quantidade de iterações, quantidade de camadas ocultas e a função de ativação nestas camadas, entre outros dados. Cada valor definido afetará diretamente o desempenho do classificador, logo, precisam ser escolhidos com um bom embasamento no assunto.

4 Resultados e Discussões

Foi construído um banco de dados extenso e com uma grande variação nas imagens de cada algarismo. O banco contava com cem imagens para cada dígito, obtendo um total de mil dados de treinamento. A eficiência da rede construída atingiu 96,33%, um valor satisfatório, se baseando em trabalhos de reconhecimento de dígitos [4]. A tabela 1 mostra a matriz de confusão com os valores dados em porcentagem, tendo em vista que as linhas representam os rótulos conhecidos, enquanto as colunas representam os rótulos previstos.

Tabela 1: Matriz de confusão da rede neural criada

Dígitos	0	1	2	3	4	5	6	7	8	9
0	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1	0%	93%	0%	0%	3%	0%	0%	0%	3%	0%
2	0%	0%	93%	0%	0%	0%	0%	7%	0%	0%
3	0%	0%	0%	90%	0%	7%	0%	3%	0%	0%
4	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
5	0%	0%	0%	0%	0%	97%	0%	0%	0%	0%
6	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%
7	0%	0%	3%	0%	0%	0%	0%	97%	0%	0%
8	0%	3%	0%	0%	0%	0%	0%	0%	97%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	3%	97%

Um arquivo *guide* foi criado e confeccionado no intuito de dinamizar o projeto, tornando-o aplicável e fácil de ser verificado, conforme figura 7 a) e b).

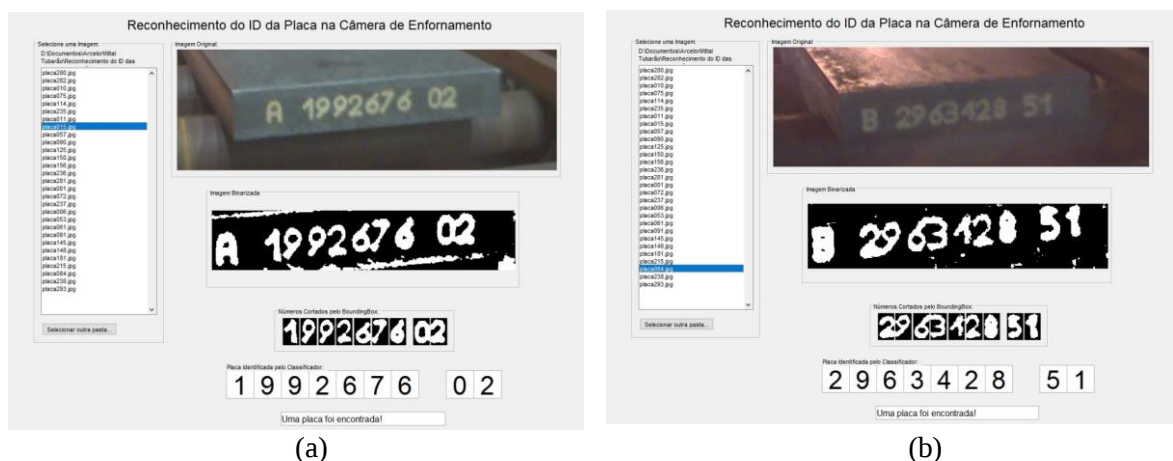


Figura 7: Classificador reconhecendo a Placa: (a) iluminação normal (b) iluminação ruidosa

5 Conclusão

Atualmente, o Deep Learning vem sendo utilizado como solução para diversos problemas de visão computacional, obtendo resultados significativos, com altas taxas de acertos.

Visando uma forma de solucionar a baixa eficiência do atual sistema de reconhecimento do ID das placas presentes no LTQ de uma Siderurgia, o presente trabalho propôs um tratamento prévio da imagem, seguido de um novo classificador. Ao aplicar os filtros de mediana e gaussiano, as principais informações da placa não foram perdidas e os ruídos foram atenuados, tornando-se possível o corte dos dígitos do ID escrito na placa via codificação em *Matlab*, mostrando ser um ótimo caminho para vencer a falta de uniformidade das imagens colhidas.

Com uma rede neural convolucional *Deep Learning* para extrair as características e classificar os dígitos, uma eficiência de 96,33% foi alcançada na rotulação das imagens de entrada, superando substancialmente o atual sistema que utiliza *Tesseract* para reconhecer os dígitos.

Portanto, com a análise das respostas obtidas, o projeto mostrou ser uma alternativa viável para solucionar o problema ocorrente na Siderurgia, obtendo resultados próximos a 100%.

A metodologia aplicada, possui potencial de ser aplicada e a mesma se encontra em estado de análise de implementação no processo real.

Referências

- [1] H.D. Hubel and T.N. Wiesel, Receptive Fields of Single neurones in the Cat's Striate Cortex.' Journal of Physiology. Vol 148, pp. 574-591, (1959).
- [2] R. Jain, R. Kasturi, and G.B. Schunck, Machine Vision. McGraw-Hill, first Edition, (1995).
- [3] A. Krizhevsky, I. Sutskever, and G. Hinton, ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25 1090–1098 (2012).
- [4] Y. LeCun, Backpropagation applied to handwritten zip code recognition. Neural Comput., 1(4):541–551, December (1989).
- [5] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, November (1998).
- [6] H. Pedrini, Apostila de Processamento de Imagens. Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações. Departamento de Informática-UFPR, (2004).
- [7] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach. [S.l.]: Prentice Hall, 3 ed., (2013).
- [8] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. Lecun. Pedestrian detection with unsupervised multi-stage feature learning, in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2013).
- [9] C. Szegedy, W. ,Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. and Rabinovich, Going deeper with convolutions, in CVPR, (2015).
- [10] "Tesseract Open Source OCR Engine (main repository)," Disponível em: <https://github.com/tesseract-ocr/>. Acesso 11 de Ago. 2017.
- [11] J. White, and G. Rohrer, Image thresholding for optical character recognition and other applications requiring character image extraction. IBM J Res Dev., 27 (4): 400-411, (1983).
- [12] L. Wu, B. Yao, Z. Peng, and Y. Guan, An adaptive threshold algorithm for sensor fault based on the grey theory. Advances in Mechanical Engineering; 9(2): 1–7, (2017).