ConBRepro
X CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO
EVENTO ON-LINE
02 a 04 de dezembro 2020

Reconhecimento de dígitos numéricos manuscritos baseado em segmentação de imagens e Redes Neurais Convolucionais

João Vargas Neto

Campus Jandaia do Sul - UFPR

Jorge Luiz Dos Santos Canuto

Campus Jandaia do Sul - UFPR

Anne Gabrielle Kessa Piai

Campus Jandaia do Sul – UFPR

Edwin Vladimir Cardoza Galdamez

Programa de Pós-graduação em Engenharia de Produção - UEM

Rodrigo Clemente Thom de Souza

Campus Jandaia do Sul - UFPR

Programa de Pós-graduação em Engenharia de Produção - UEM

Resumo: A avaliação de crianças em fase de alfabetização trata-se de um processo em que o professor avalia a aquisição tanto da base alfabética, quanto numérica. Este processo pode ser demorado e de difícil realização. A fim de automatizar esta tarefa docente, o presente artigo propõe a aplicação de uma solução de processamento digital de imagens envolvendo segmentação e reconhecimento de dígitos numéricos manuscritos. Para segmentação foram aplicados histogramas, limiarização de intensidade e detecção de bordas. Para classificação são comparadas diferentes arquiteturas de Redes Neurais Convolucionais (CNN). O conjunto de dados de dígitos manuscritos trata-se do benchmark Modified National of Standards and Technology (MNIST). Os resultados indicam que as CNNs foram acuradas para a tarefa a que foram submetidas, com especial destaque para a arquitetura VGG16.

Palavras-chave: Segmentação de imagem, Redes Neurais Convolucionais, Classificação, MNIST.

Handwritten numeric digits recognition based on images segmentation and Convolutional Neural Networks

Abstract: The assessment of children in the literacy phase is a process in which the teacher assesses the acquisition of both the alphabetical and numerical basis. This process can be time-consuming and difficult to perform. In order to automate this teaching task, this article proposes the application of a digital image processing solution involving segmentation and recognition of handwritten numeric digits. For segmentation, histograms, intensity threshold and edge detection were applied. For classification, different architectures of Convolutional Neural Networks (CNN) are compared. The handwritten digit data set is the Modified National of Standards and Technology (MNIST) benchmark. The results indicate that the CNNs were accurate for the task to which they were submitted, with special emphasis on the VGG16 architecture.

Keywords: Image segmentation, Convolutional Neural Networks, Classification, MNIST.

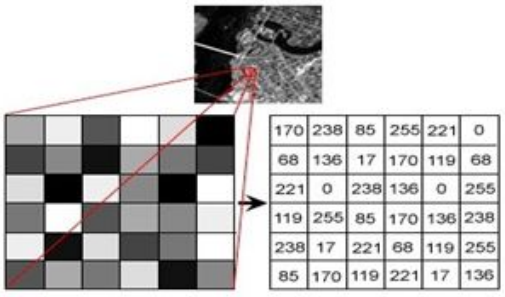
1. Introdução

A avaliação de crianças em fase de alfabetização trata-se de um processo em que o professor avalia a aquisição tanto da base alfabética, quanto numérica. Este processo pode ser demorado e de difícil realização. A fim de automatizar esta tarefa docente, o presente artigo propõe a aplicação de uma solução de processamento digital de imagens envolvendo segmentação e reconhecimento de dígitos numéricos manuscritos. Para tanto, os números escritos pelos alunos poderiam ser digitalizados na forma de imagens e, posteriormente, aplicados os algoritmos para segmentação de cada dígito e, em seguida, os algoritmos para reconhecimento (classificação) de cada dígito isoladamente.

Primeiramente, é importante ter alguns conceitos em mente sobre o funcionamento dos algoritmos para processamento digital de imagens. Uma imagem, a grosso modo, é “vista” pelo computador em preto e branco, mas na verdade ela é representada em níveis de uma escala de cinza, na qual estes níveis variam entre as tonalidades do preto (menor intensidade, 0) até e o branco (maior intensidade, 255). Estes valores são para uma imagem de tamanho 16x16 com 256 pixels de resolução (variação entre 0 e 255), e isto é representado em uma matriz bidimensional onde cada variação de tonalidade representa um valor de um pixel específico.

A Figura 1 apresenta uma imagem em escalas de cinza (*grayscale*).

Figura 1 – Imagem em escalas de cinza.



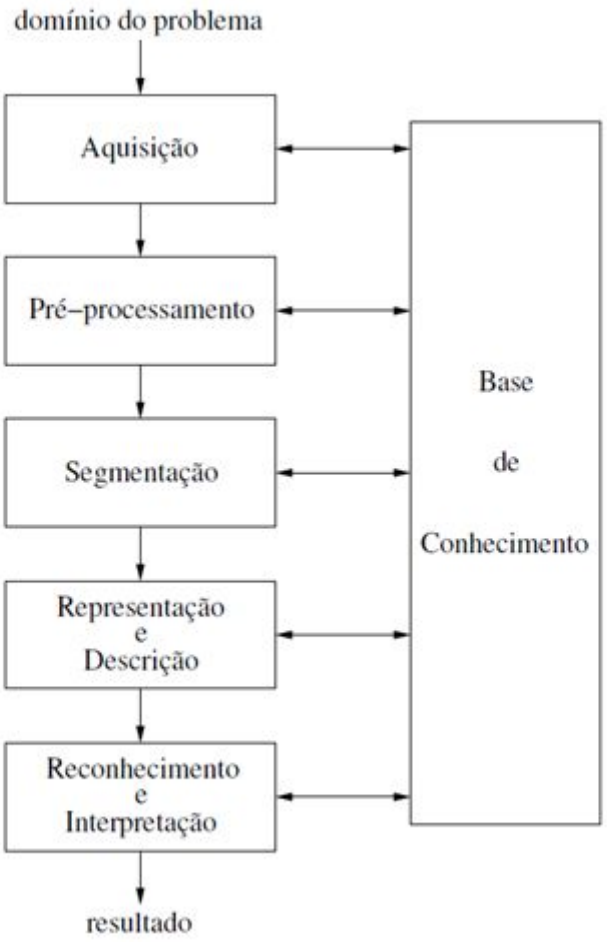
Fonte: Clappis (2009).

O processamento digital de imagens trata-se de um conjunto de processos interligados que se inicia na captura da imagem e perpassa até ser digitalizada para então ser representada de maneira adequada em visão computacional.

Também é realizado o pré-processamento, a filtragem de ruídos, correção de distorções geométricas (extração de bordas, texturas, vizinhanças e movimento), para então segmentar a imagem e classificá-la.

A Figura 1 apresenta um exemplo de procedimento de processamento digital de imagens.

Figura 2 - Tarefas no processamento digital de imagens.



Fonte: Pedrini (2020)

2. Convolutional Neural Networks (CNN)

As *CNNs* são um dos vários tipos de redes neurais, e elas possuem três camadas, a convolucional, a de agrupamento e a camada totalmente conectada. Essas camadas, respectivamente, extraem características da imagem, transforma os valores de entrada em valores representativos junto com a redução das dimensões espaciais para então identificar padrões realizar classificações.

As *CNNs* possuem diferentes tipos de arquiteturas, onde cada tipo possui alguma mudança/alteração específica em seu escopo, em relação a outra. Essas mudanças influenciam diretamente no resultado final da análise.

Diversos trabalhos vem sendo publicados ao longo do últimos anos sobre a aplicação de CNNs para classificação das imagens do conjunto *MNIST* (LECUN, 1998), dentre eles se podem destacar estudos pioneiros como o de KUSSUL e BAIDYK (2004), e trabalhos mais recentes, como os de AHLAWAT e CHOUDHARY (2020), KAYUMOV et al. (2020) e GUPTA e BAG (2021).

Neste trabalho serão apresentados 6 tipos de arquiteturas de redes neurais (NN), sedo elas: DenseNet121, DenseNet201, ResNet50, MobileNet, VGG16 e VGG19. A

fim de testá-las e avaliá-las para descobrir qual seria a mais adequada, levando em consideração aquela que possuir a maior acurácia.

3. Materiais e Métodos

3.1 Conjunto de dados

Os dados utilizados são provenientes da base de dados *MNIST* que se tornou popular e comumente usada para testar diferentes algoritmos de *ML* e *DL*, principalmente as *CNNs.* O *MNIST* contém imagens de dígitos manuscritos do número 0 até o número 9 com dimensões 28x28. Além disso, possui um total de 70.000 imagens rotuladas, com 60.000 imagens para treinamento e 10.000 imagens para teste.

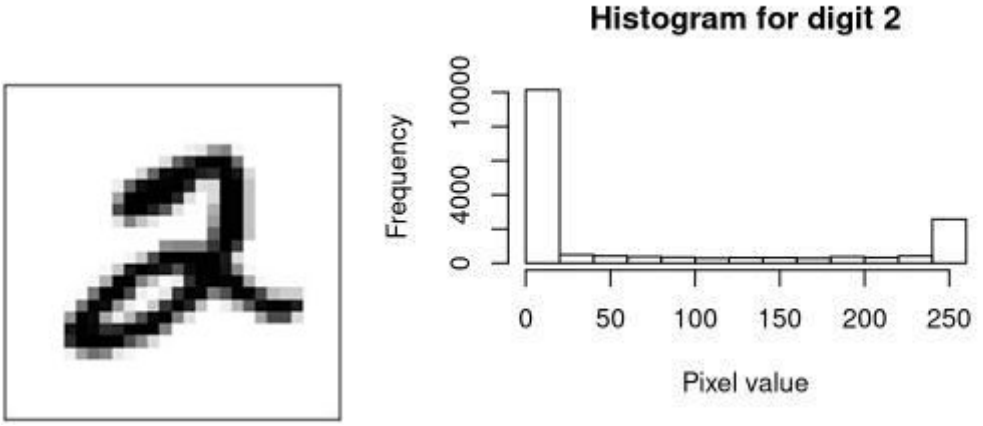
3.2 Metodologia

Para o presente trabalho utilizamos o *Google Colaboratory,* que possui processamento em nuvem, suporte de aceleração de hardware, tanto para CPU quanto GPU. O Colab usa a linguagem Python e conta com várias bibliotecas disponíveis inclusive bibliotecas Keras.

Para a etapa de segmentação das imagens for gerados histogramas, aplicado o limiar de intensidade e a detecção de bordas.

Também conhecidos como diagrama de dispersão de frequência, os histogramas deste artigo apresentam as distribuições das variações das tonalidades de cada pixel da imagem.

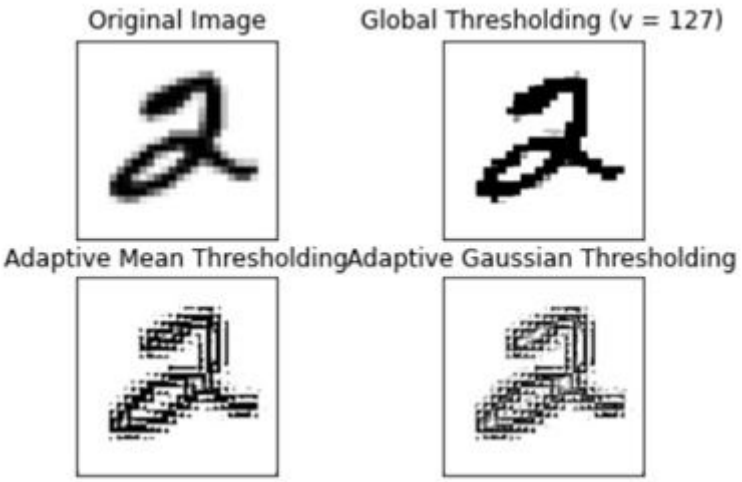
Figura 3 – Exemplo de histograma de uma das imagens de dígitos manuscritos.



Fonte: Cichy (2019)

Quanto ao limiar de intensidade, é aplicado para se analisar a similaridade de níveis de cinza a fim de extrair objetos específicos em função de um limiar T que separa os agrupamentos de escala de cinza da imagem. A intensidade é quantidade de preto no branco.

Figura 4 – Exemplo de limiar de intensidade de uma das imagens de dígitos manuscritos.

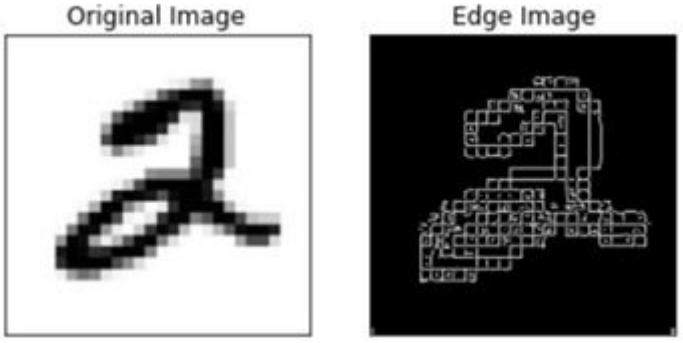


Fonte: Os autores.

A detecção de borda é usada para identificar os elementos chaves que separam a figura do fundo da imagem. Em outras palavras a borda é um tipo de fronteira entre as regiões da imagem (número e fundo).

Para realizar a detecção são aplicados filtros usando derivadas, juntamente com as máscaras convolucionais.

Figura 5 - Exemplo de detecção de bordas de uma das imagens de dígitos manuscritos.



Fonte: Os autores.

As arquiteturas de *CNN* utilizadas para reconhecimento dos dígitos após a aplicação da segmentação foram: DenseNet121 (HUANG *et al*., 2017), ResNet50 (HE *et al*., 2016), MobileNet (HOWARD *et al*., 2017), VGG16 (SIMONYAN e ZISSERMAN, 2014), VGG19 (SIMONYAN e ZISSERMAN, 2014) e DenseNet201 (HUANG *et al*., 2017). Os treinamentos de todas as arquiteturas foram realizados com 60.000 imagens em 10 épocas e os testes ocorreram com 10.000 imagens. As métricas

para avaliação comparativa de desempenho das arquiteturas foram: *Accuracy* (Eq. 1), *Precision* (Eq. 2), *Recall* (Eq. 3) e *F1-Score* (Eq. 4).

$$Acc = \frac{Totaldeclassificações}{Quantidadedeacertos} \quad (1)$$

$$Prec = \frac{VP}{(VP+FP)} \quad (2)$$

$$Rec = \frac{VP}{(VP+FN)} \quad (3)$$

$$F1 = \frac{2*(precisão*revocação)}{(precisão+revocação)} \quad (4)$$

Utilizamos uma abordagem de transferência de aprendizagem que inicializa os pesos do modelo treinado em um conjunto de dados maior, nesse caso utilizamos os pesos do ImageNet. Cada arquitetura foi utilizada na prática como extrator de recursos fixo, ou seja, congelamos a parte inferior do modelo e treinamos apenas as camadas totalmente conectadas.

Também foi usado o redimensionamento de imagem para adaptar as dimensões das imagens do conjunto *MNIST*, que é 28x28, a entrada das *CNNs,* que possuem como menor entrada possível a forma 32x32.

4. Resultados e discussão

O desempenho geral de um modelo é dado pela acurácia. As acurácias obtidas na etapa experimental são apresentadas na Tabela 1.

Tabela 1-Acurácias.

Arquitetura	Acurácia
DenseNet121	0.97
DenseNet201	0.90
ResNet50	0.95
MobileNet	0.97
VGG16	0.98
VGG19	0.87

Fonte: Os autores.

Em termos de acurácia observamos que a arquitetura VGG16 apresentou o melhor resultado, inclusive quando comparamos com a sua variante VGG19. De acordo com os resultados, percebemos nem sempre o aumento da complexidade da rede por meio da adição de camadas resulta em um melhor desempenho.

As tabelas 2, 3, 4, 5, 6 e 7 mostram os resultados para as demais métricas.

Tabela 2-Resultado DenseNet121

Classe	Precisão	Revocação	Pontuação f1	Quantidade de imagens
0	0.98	0.99	0.98	980
1	0.99	0.99	0.99	1135
2	0.96	0.97	0.97	1032
3	0.97	0.98	0.98	1010
4	0.97	0.98	0.97	982
5	0.98	0.97	0.98	892
6	0.98	0.98	0.98	958
7	0.97	0.97	0.97	1028
8	0.97	0.97	0.97	974
9	0.98	0.94	0.96	1009
macro avg	0.97	0.97	0.97	10000
weighted avg	0.97	0.97	0.97	10000

Fonte: Os autores.

Tabela 3-Resultado DenseNet201

Classe	Precisão	Revocação	Pontuação f1	Quantidade de imagens
0	0.95	0.97	0.96	980
1	0.95	0.98	0.97	1135
2	0.85	0.82	0.83	1032
3	0.91	0.90	0.90	1010
4	0.89	0.92	0.91	982
5	0.84	0.84	0.84	892
6	0.94	0.95	0.95	958
7	0.92	0.87	0.89	1028
8	0.89	0.91	0.90	974
9	0.89	0.89	0.89	1009
macro avg	0.90	0.90	0.90	10000
weighted avg	0.90	0.90	0.90	10000

Fonte: Os autores.

Tabela 4-Resultado MobileNet.

Classe	Precisão	Revocação	Pontuação f1	Quantidade de imagens
0	0.96	0.98	0.97	980
1	0.97	0.98	0.98	1135
2	0.94	0.94	0.94	1032
3	0.94	0.97	0.95	1010
4	0.95	0.95	0.95	982
5	0.95	0.94	0.94	892
6	0.96	0.95	0.96	958
7	0.94	0.96	0.95	1028
8	0.92	0.94	0.93	974
9	0.97	0.90	0.93	1009
macro avg	0.95	0.95	0.95	10000
weighted avg	0.95	0.95	0.95	10000

Fonte: Os autores.

Tabela 5-Resultado ResNet50

Classe	Precisão	Revocação	Pontuação f1	Quantidade de imagens

0	0.97	0.99	0.98	980
1	0.98	0.99	0.98	1135
2	0.97	0.97	0.97	1032
3	0.96	0.98	0.97	1010
4	0.97	0.97	0.97	982
5	0.97	0.97	0.97	892
6	0.97	0.98	0.97	958
7	0.97	0.95	0.96	1028
8	0.96	0.96	0.96	974
9	0.96	0.95	0.95	1009
macro avg	0.97	0.97	0.97	10000
weighted avg	0.97	97	0.97	10000

Fonte: Os autores.

Tabela 6-Resultado VGG16

Classe	Precisão	Revocação	Pontuação f1	Quantidade de imagens
0	0.99	0.99	0.99	980
1	0.99	0.99	0.99	1135
2	0.99	0.98	0.98	1032
3	0.99	0.98	0.98	1010
4	0.98	0.99	0.98	982
5	0.98	0.99	0.98	892
6	0.99	0.98	0.99	958
7	0.98	0.98	0.98	1028
8	0.98	0.98	0.98	974
9	0.98	0.97	0.97	1009
macro avg	0.98	0.98	0.98	10000
weighted avg	0.98	0.98	0.98	10000

Fonte: Os autores.

Tabela 7-Resultado VGG19

Classe	Precisão	Revocação	Pontuação f1	Quantidade de imagens
0	0.91	0.94	0.93	980
1	0.94	0.97	0.96	1135
2	0.75	0.79	0.77	1032
3	0.87	0.87	0.87	1010
4	0.85	0.90	0.87	982
5	0.82	0.75	0.78	892
6	0.94	0.94	0.94	958
7	0.89	0.74	0.81	1028
8	0.84	0.88	0.86	974
9	0.85	0.85	0.85	1009
macro avg	0.87	0.86	0.86	10000
weighted avg	0.87	0.87	0.86	10000

Fonte: Os autores.

A arquitetura VGG16 novamente apresenta os melhores resultados em termos de precisão, revocação e pontuação f1. A superioridade da VGG16 pode ser explicada por uma melhor extração automática de recursos das imagens contidas na tarefa em questão.

5. Considerações finais

No presente trabalho foram esclarecidos conceitos essenciais para o entendimento e funcionamento do processamento digital de imagens em conjunto de dados MNIST, a fim de testá-las para descobrir qual apresenta melhores resultados para a tarefa de classificação de dígitos manuscritos feito por crianças em fase de alfabetização.

Os resultados indicam que as CNNs foram acuradas para a tarefa a que foram submetidas. A partir da comparação entre as arquiteturas ficou explícito que a VGG16 seria a mais adequada para a tarefa de classificação, isto em função de seu aprendizado mais rápido em relação as demais arquiteturas. Observamos também que o aumento da complexidade do modelo por meio da adição de camadas nem sempre resulta em um melhor desempenho. Outras CNNs com bom desempenho foram as arquiteturas DenseNet121 e ResNet50.

Referências

AHLAWAT, S., CHOUDHARY, A. Hybrid CNN-SVM Classifier for Handwritten Digit Recognition. **Procedia Computer Science**, pp. 2554-2560, Volume 167, 2020.

CICHY, T. **MNIST R Deep Learning.** Dísponivel em: <https://tomaszcichy.com/data/r_deep/mnist#5_references_and_links>Acessado em 25 set. 2020.

CLAPPIS, A. M. **Uma introdução as redes neurais convolucionais utilizando o Keras** Disponível em: <https://medium.com/data-hackers/uma-introdu%C3%A7%C3%A3o-as-redes-neurais-convolucionais-utilizando-o-keras-41ee8dcc033e>Acesso em: 21 set. 2020.

GONZALEZ, R. C; WOODS, R. E. **Processamento Digital de Imagens.** Ed.Pearson, 2011.

GUPTA, D., BAG, S. CNN-based multilingual handwritten numeral recognition: A fusion-free approach. **Expert Systems with Applications**, Volume 165, 113784, 2021.

HE, K; ZHANG, X; REN, S; SUN, J. **Deep Residual Learning for Image Recognition.** Disponível em: <https://arxiv.org/pdf/1512.03385.pdf> Acesso em 25 set. 2020.

HOWARD, A. G *et al*. **MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.** Disponível em: <https://arxiv.org/pdf/1704.04861.pdf> Acesso em 25 set. 2020.

HUANG, G; LIU. Z; MAATEN, L. **Densely Connected Convolutional Networks.** Disponivel em: <https://arxiv.org/pdf/1608.06993.pdf> Acesso em 25 set.2020.

KAYUMOV, Z., TUMAKOV, D., MOSIN, S. Hierarchical Convolutional Neural Network for Handwritten Digits Recognition. **Procedia Computer Science**, pp. 1927-1934, Volume 171, 2020.

KUSSUL, R., BAIDYK, T. Improved method of handwritten digit recognition tested on MNIST database. **Image and Vision Computing**, pp. 971-981, Volume 22, Issue 12, 2004.

LECUN, Yann. The MNIST database of handwritten digits. **http://yann.lecun.com/exdb/mnist/**, 1998.

PEDRINI, H. **Introdução ao Processamento Digital de Imagem MC920 / MO443.** Disponível em: <https://www.ic.unicamp.br/~helio/disciplinas/MC920/aula_introducao.pdf> Acesso em 25 set. 2020.

SIMONYAN, K; ZISSERMAN, A. **Very deep convolutional networks for large-scale image recognition.** Disponível em: <https://arxiv.org/pdf/1409.1556.pdf> Acesso em 25 set. 2020.