

Método para Ampliação de Imagens Baseado em Rede Neural

Lázaro Ribeiro Monteiro Júnior¹, Silvana dos Santos Lima Morais¹

¹Programa de Pós-Graduação em Ciência da Computação (PPgCC) – Universidade do Estado do Rio Grande do Norte (UERN) – CEP: 59600-970, Mossoró, RN, Brasil & Universidade Federal Rural do Semi-Árido (UFERSA) – CEP: 59625-900, Mossoró, RN, Brasil.

lazaromjr@outlook.com, silvanasali88@gmail.com

Abstract. *The single image upscaling is a well-known problem from Digital Image Processing area. It consists at increase an image resolution's for better viewing. However, this process creates empty spaces between the pixels from the original image and fulfill these empty spaces is not a simple task. The most stablished treatments for this problem are based in interpolation. This paper describes a method based on neural network.*

Resumo. *A ampliação de imagens é um problema conhecido da área de Processamento Digital de Imagens. Ele consiste em aumentar a resolução da imagem para melhor visualização. No entanto, esse processo cria espaços vazios entre os pixels da imagem original e preencher esses espaços vazios não é uma tarefa simples. Os tratamentos estabelecidos para esse problema são baseados em interpolação. Este artigo descreve um método baseado em rede neural.*

1. Introdução

A ampliação de imagens é um problema bem estabelecido na área de processamento digital de imagens (PDI). Ele consiste em aumentar as dimensões da imagem para melhorar sua visualização, observar detalhes ou processá-la. O processo de ampliação envolve a inclusão de novos *pixels* na imagem. Contudo, definir os valores desses *pixels* não é uma tarefa trivial, uma vez que eles são desconhecidos. Os *pixels* são os menores elementos que formam uma figura, seu nome é uma abreviação de *picture elements* (elementos de figura) (Gonzalez, 2010).

As técnicas mais comumente utilizadas consistem em repetir as cores dos pixels adjacentes ou utilizar interpolação para estimá-las. Mesmo assim, as imagens perdem definição e adquirem aspecto serrilhado ou embaçado, como vemos na Figura 1.



Figura 1. Aspecto serrilhado após a ampliação

Fonte: Autoria própria

As técnicas de ampliação mais utilizadas são: vizinho mais próximo, interpolação bilinear e interpolação bicúbica. A interpolação é uma importante ferramenta empregada para aferir valores em pontos desconhecidos a partir de uma amostra de dados conhecidos (Gonzalez, 2010). Existem diversos outros métodos, porém, a relação custo computacional/qualidade dos resultados levam estes três métodos a serem mais utilizados.

Na interpolação pelo vizinho mais próximo (replicação) a cor de um *pixel* é determinada como sendo a cor do *pixel* definido mais próximo. A interpolação bilinear utiliza os 4 pixels vizinhos para calcular a cor do novo *pixel* através de um cálculo de média ponderada. A interpolação bicúbica funciona de forma semelhante à bilinear, porém utiliza uma vizinhança de 16 *pixels* em seus cálculos (Gonzalez, 2010).

Com o avanço da tecnologia a presença de agentes inteligentes que utilizam técnicas de aprendizagem de máquina se tornam cada vez mais presentes em nosso cotidiano. Desta forma, o presente estudo tem como objetivo desenvolver um método de ampliação de imagens baseado em redes neurais para estimar os valores dos novos *pixels*.

O uso de redes neurais artificiais no processo de ampliação de imagens busca substituir a interpolação no processo, utilizando a rede neural para prever os valores dos *pixels*. O uso de aprendizagem de máquina tem o potencial de trazer novos resultados para este problema e pode criar uma variedade de estudos na área, uma vez que as redes neurais possuem muitas especificidades.

Este documento é organizado da seguinte maneira: a Seção 2 apresenta os trabalhos relacionados; a Seção 3 descreve o método desenvolvido; a Seção 4 apresenta os testes realizados e os resultados obtidos; por fim, a Seção 5 apresenta as considerações finais.

2. Trabalhos Relacionados

Nesta seção são apresentados os trabalhos relacionados ao presente estudo.

2.1. Single Image Super-Resolution Using Back-Propagation Neural Networks (Super Resolução de Imagem Usando Redes Neurais de Propagação)

Existem vários algoritmos matemáticos para melhorar a imagem colorida, como o do vizinho mais próximo, bicúbico e bilinear. Este artigo faz uma comparação de desempenho entre esses três algoritmos e verificou-se que o bicúbico tem o melhor desempenho em termos de similaridade estrutural e tempo de execução. Um método bicúbico com método de Redes Neurais Artificiais (RNA) de retro propagação foi proposto para melhorar os resultados.

O bicúbico com RNA mostra SSIM (Structural Similarity) 6,5% maior; o PSNR (Peak Signal-to-Noise Ratio) 6,9% maior; o SNR (Signal-to-Noise Ratio) 8,7% maior e o MSE (Mean Squared Error) 30,23% menor. Os resultados do bicúbico com RNA também são comparados com as técnicas de super-resolução de última geração como SRCNN (Super-Resolution Convolutional Neural Network). O bicúbico com RNA produz SSIM 1,48% maior e o PSNR 3,44% maior que o SRCNN.

2.2. A New Approach for Upscaling Document Images for Improving Their Quality (Uma Nova Abordagem para Ampliar Imagens de Documentos para Melhorar sua Qualidade)

Um dos problemas enfrentados pelos softwares de reconhecimento óptico de caracteres (OCR) é que as imagens dos documentos de entrada não são de boa qualidade. O objetivo dessa pesquisa é demonstrar um método para gerar uma imagem de documento de alta resolução, dada uma imagem de baixa resolução.

Foi proposto um novo método para melhorar a resolução espacial das imagens dos documentos. Construiu-se um modelo baseado em rede neural profunda que utiliza os métodos tradicionais de interpolação, obteve-se os melhores recursos e reconstruiu-se uma imagem de alta resolução a partir desses recursos. Isso foi conseguido usando uma rede neural convolucional (CNN).

A técnica utilizada nesse trabalho foi denominada de fusão não linear de múltiplas interpolações (NFMI). O método NFMI garante que o modelo aprenda apenas os melhores recursos que podem ser extraídos de todas as técnicas de interpolação combinadas. O uso de métodos tradicionais de interpolação garante que a técnica NFMI não seja computacionalmente cara.

Os resultados nas imagens de teste mostram uma melhoria relativa de 54% na precisão do reconhecimento de palavras pelo OCR sobre a melhor técnica de interpolação para duplicar a resolução espacial e 33% para quadruplicar a resolução.

2.3. Zoom to Learn, Learn to Zoom (Ampliar para Aprender, Aprender para Ampliar)

Esse artigo mostra que, ao utilizar o aprendizado de máquina no zoom digital para imagens, é útil utilizar dados reais do sensor (RAW) para o treinamento. Os métodos de super-resolução baseados em aprendizado existentes não usam dados reais do sensor, mas operam em imagens RGB (Red, Green e Blue).

Na prática, essas abordagens procedem em perda de detalhes e precisão na saída com zoom digital ao aumentar o zoom em regiões distantes da imagem. A principal barreira para o uso de dados reais do sensor para treinamento é a falta de imagens de alta resolução.

Nesse trabalho foi mostrado como conseguir os dados reais das imagens por meio de zoom óptico e como contribuir com um conjunto de dados, SR-RAW, para o zoom computacional. Como resultado, foi demonstrado a eficácia do uso de dados coletados no sensor para zoom computacional. As imagens são ampliadas a partir de dados brutos do sensor através de um modelo de aprendizado profundo que executa ISP (Image Signal Processor) em um conjunto com super-resolução. A abordagem absorve o sinal dos dados brutos e produz resultados de alta qualidade comparada aos modelos treinados em imagens RGB processadas ou em dados de sensores sintéticos.

2.4. Scale-Iterative Upscaling Network for Image Deblurring (Rede de Ampliação em Escala Iterativa para Desembaçar Imagem)

Os métodos baseados em aprendizado de máquina para desembaçar imagens são eficientes para lidar com imagens desfocadas do mundo real, cujo desfoque pode ser causado por várias distorções combinadas. No entanto, as arquiteturas multiníveis existentes não conseguem ajustar imagens de vários cenários.

Nesse artigo foi proposto uma rede de aumento de escala iterativa (Scale-Iterative Upscaling Network - SIUN) que restaura imagens nítidas de maneira iterativa. Ele não apenas é capaz de preservar o compartilhamento de pesos entre escalas, mas também é mais flexível ao treinar e prever com diferentes iterações para ajustar imagens diferentes.

Especificamente, foi exposta a estrutura de super-resolução em vez da camada de aumento entre duas escalas consecutivas para restaurar uma imagem detalhada. Além disso, explorou-se diferentes estratégias de aprendizagem para treinamento e previsão da rede e introduziu-se uma estratégia amplamente aplicável para tornar o SIUN compatível com diferentes cenários, incluindo texto e face.

Resultados experimentais em conjuntos de dados de referência e imagens desfocadas reais mostram que esse método pode produzir melhores resultados do que os métodos de ponta.

3. Metodologia

O procedimento para expandir uma imagem, aumentando seu tamanho, consiste em “separar” os *pixels* da imagem, criando entre eles lacunas que devem ser posteriormente preenchidas. A Figura 2 ilustra o processo de ampliação de uma imagem de 5×5 *pixels* por fator 3, ou seja, a imagem terá seu tamanho aumentado em 3 vezes.

Em seguida, as lacunas criadas são preenchidas com valores estimados, uma vez que essas posições não existiam previamente na imagem. Esse passo define a qualidade da imagem resultante, podendo deixá-la embaçada ou serrilhada, deteriorando sua qualidade. A segunda etapa é ilustrada na Figura 3.

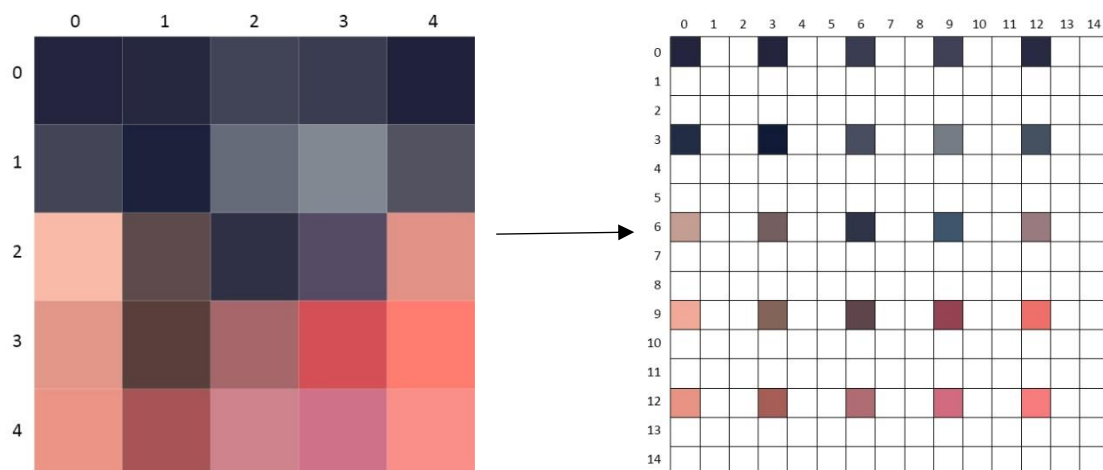


Figura 2. 1ª etapa do processo de ampliação

Fonte: Autoria própria

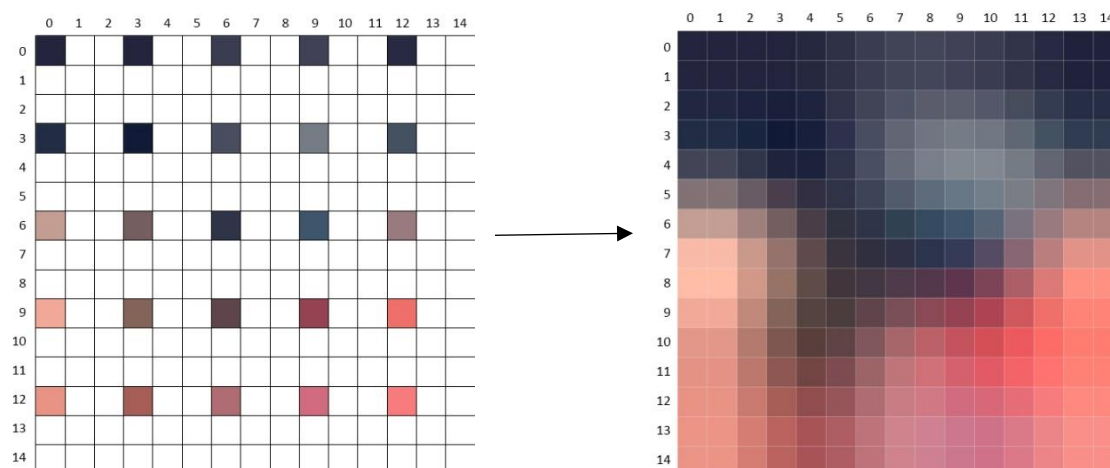


Figura 3. 2ª etapa do processo de ampliação

Fonte: Autoria própria

3.1. Rede Neural Artificial

O neurônio artificial busca simular o funcionamento de células cerebrais, que são responsáveis por enviar e receber impulsos nervosos. Cada célula é composta por três divisões com funções específicas e complementares: corpo, dendritos e axônio. O cérebro é constituído de bilhões de neurônios e essa troca de sinais pode se reproduzir em diversas camadas de neurônios (Haykin, 2001).

O neurônio artificial, de forma análoga ao natural, é constituído por três elementos básicos: um conjunto de n conexões de entrada; um somador (Σ) para guardar os sinais de entrada; e uma função de ativação (φ) que restringe a amplitude do sinal de saída (y) a um valor fixo (Haykin, 2001). A Figura 4 apresenta a estrutura de um neurônio artificial.

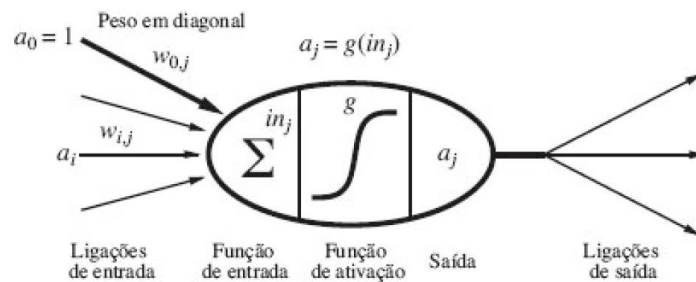


Figura 4. Modelo de um neurônio artificial de McCulloch e Pitts

Fonte: Russell, 2010

A combinação de diversos neurônios forma uma rede neural artificial, um método de inteligência artificial baseado no modelo de funcionamento neuronal do cérebro humano (Haykin, 2001). O *perceptron* é uma rede neural de única camada e possui a capacidade de categorizar padrões linearmente separáveis, ou seja, caso um *perceptron* seja criado com um único neurônio, esse será apto de separar padrões somente entre duas classes. Contudo, com o progresso da quantidade de neurônios, a quantidade de classes separáveis crescerá relativamente (Haykin, 2001).

Há também o *perceptron* multicamadas, utilizado quando o problema abordado não admite uma solução linear. Uma RNA *Multi Layer Perceptron (MLP)* é subdividida em camadas: sendo elas camada de entrada, camada(s) intermediária(s) ou ocultas(s) e camada de saída. A camada de entrada entrega as informações de entrada para a(s) camada(s) oculta(s) da rede. As camadas ocultas são camadas intermediárias, que têm como tarefa separar as camadas de entrada e de saída. Os neurônios de uma camada estão interligados somente aos neurônios da camada posterior, não ocorrendo retroalimentação, nem interligações entre neurônios da mesma camada. Na camada de saída, a solução do problema é alcançada. Tipicamente, as camadas são completamente conectadas (Haykin, 2001).

A rede neural artificial *multi layer perceptron* utilizada para tratar o problema abordado neste estudo é representada na Figura 5 através de um grafo orientado, onde os nós representam os neurônios e as arestas o fluxo dos sinais. O fluxo de sinais por meio da rede é executado corretamente, da esquerda para a direita, camada a camada.

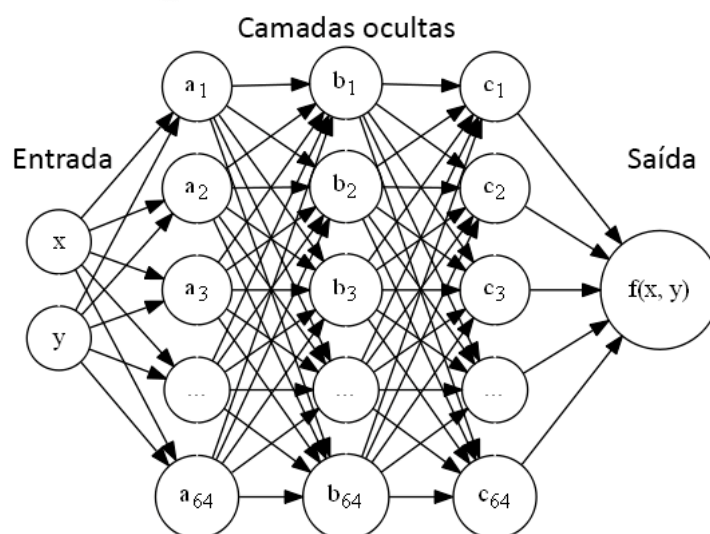


Figura 5. Rede Neural Artificial Multicamadas

Fonte: Autoria própria

A camada de entrada da rede possui dois neurônios que são alimentadas com as coordenadas (x, y) de cada *pixel*. Em seguida, os sinais são disseminados em 3 camadas ocultas, com 64 neurônios cada que utilizam função de ativação *relu*. Por fim, a última camada de apenas um neurônio e de função de ativação linear deve fornecer o valor de intensidade do *pixel* dado as entradas (x, y) . As entradas e saídas possuem valor normalizado em escala entre 0 e 1.

3.2. Treinamento

Para que uma rede neural aprenda e se adapte a um determinado problema, é preciso que passe por uma fase de treinamento. Nessa fase os pesos ligados às conexões são ajustados seguindo algumas regras definidas. Essas regras de treinamento descrevem como os pesos podem ser ajustados durante o aprendizado para aperfeiçoar o desenvolvimento da rede. Após a etapa de treinamento os parâmetros da rede permanecem estáticos (Haykin, 2001).

O aprendizado das redes neurais artificiais pode ser classificado em dois tipos, sendo eles: aprendizado supervisionado, onde a rede é treinada provendo-se um conjunto de entradas e saídas, e os pesos são ajustados conforme a diferença entre a entrada disponível e a saída almejada; e aprendizado não-supervisionado, no qual as redes recebem apenas os valores dos atributos de entrada, sendo a saída definida pela própria rede. Neste trabalho será utilizada a aprendizagem supervisionada.

No desenvolvimento do método foi utilizado o método de otimização de *Adam* e a curva de aprendizagem da rede pode ser observada na Figura 6. Observa-se que a taxa de erro diminui conforme as iterações realizadas no processo de treinamento são elevadas, o que caracteriza o processo de aprendizagem. Logo, a taxa de acerto da rede neural está diretamente relacionada com o esforço computacional realizado durante o treinamento. A rede neural proposta deve ser treinada com recortes da imagem de tamanho fixo e após treinada deve prever os novos valores que devem ser inseridos na imagem.

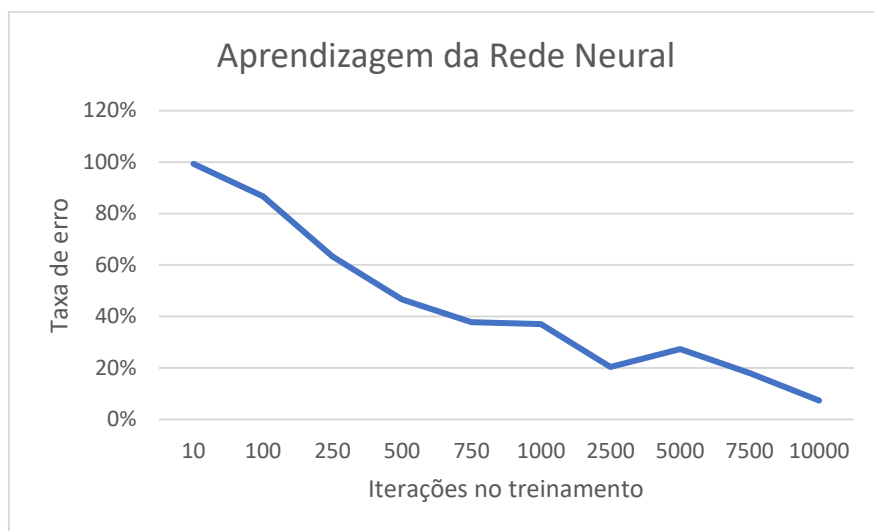


Figura 6. Curva de aprendizagem da rede neural

Fonte: Autoria própria

A implementação do método proposto para ampliação de imagens foi realizada na linguagem de programação *python*. Foram utilizadas as APIs: *Pillow*, *TensorFlow* e *Keras*. A *Pillow* é uma API que fornece recursos para manipular imagens, como abrir, extrair seus *pixels*, além de gerar, salvar e exibir as novas imagens processadas. Já as APIs *TensorFlow* e *Keras* dispõem de diversas técnicas de aprendizagem de máquina. Elas foram utilizadas para implementar a rede neural descrita da Figura 5.

4. Resultados e Discussões

Nesta seção serão descritos os testes realizados e os resultados obtidos. Os primeiros testes foram feitos a partir da Figura 7a, de tamanho 8 x 8 e criada com *pixels* aleatórios de diferentes cores. A partir desta imagem foram realizados testes, ampliando-a em 4 vezes e variando a arquitetura da rede neural MLP. As Figuras 7b, 7c, 7d, 7e, 7f, 7g apresentam os resultados obtidos nesta etapa.

É possível observar que a qualidade dos resultados evolui conforme o número de neurônios e camadas ocultas são elevados. Nas primeiras imagens é possível notar transições abruptas de intensidades de cor e falhas nas imagens. No último experimento (Figura 7g) já se observa transições suaves entre os *pixels* originais e não há presença de falhas. Sequer é possível identificar quais são os pixels originais, produzindo um efeito semelhante ao obtido na interpolação.

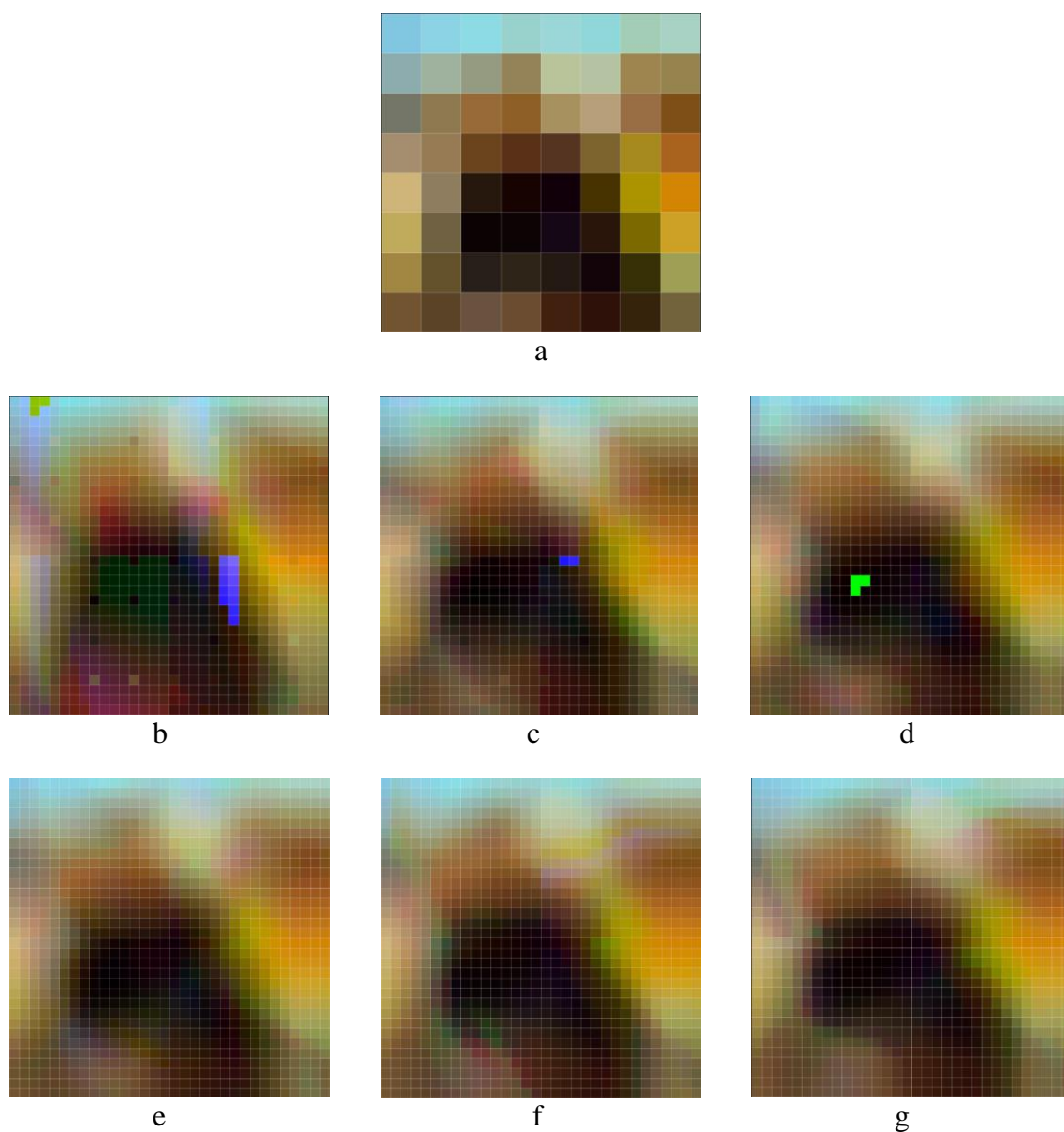


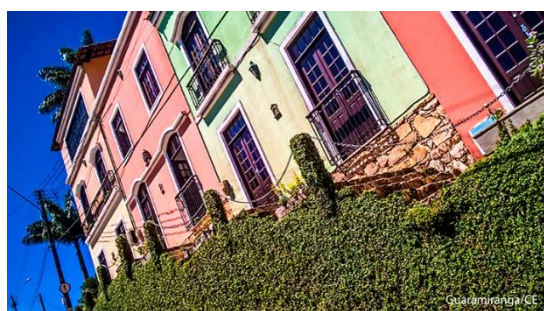
Figura 7. a) Imagem não processada b) Imagem processada com RNA de 3 camadas ocultas de 8 neurônios b) Imagem processada com RNA de 3 camadas ocultas de 32 neurônios c) Imagem processada com RNA de 3 camadas ocultas de 64 neurônios d) Imagem processada com RNA com 3 camadas ocultas de 128 neurônios e) Imagem processada com RNA de 5 camadas ocultas de 128 neurônios f) Imagem processada com RNA de 5 camadas ocultas de 256 neurônios.

As imagens exibidas na Figura 8 foram ampliadas em 2x e 4x. As imagens originais tem tamanho 712 x 400 pixels e foram reduzidas para 178 x 100 *pixels*. Em seguida foram ampliadas em 2 e 4 vezes para reestabelecer seu tamanho original através dos métodos de ampliação. As imagens resultantes são comparadas com as originais através do PSNR.

O PSNR é utilizado para definir a relação entre a energia máxima de um sinal (intensidade do pixel) e o ruído que afeta sua representação (distorção). Dessa forma é possível comparar qual imagem ampliada possui menos ruídos em relação à original.

O PSNR é expresso em escala logarítmica e utiliza como unidade o decibel (db). A Equação 1 descreve o cálculo realizado, onde MAX_I é o valor de pico do sinal, no caso, do *pixel* e MSE é o erro quadrado médio. Logo, quanto maior o valor do PSNR, menos ruído a imagem possui.

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad \text{Equação 1}$$



a



b



c

Figura 8. Imagens utilizadas nos testes

Fonte: Autoria própria

Na Tabelas 1 e 2 são apresentados os valores de PSNR dos testes realizados.

Tabela 1. PSNR das imagens ampliadas 2x

	Vizinho mais próximo	Bilinear	Bicúbico	Zoom com Rede Neural
Imagem a	16,573872	16,372977	17,695625	14,512599
Imagem b	24,490696	24,550793	26,498424	19,377241
Imagem c	25,964929	25,828712	27,242227	22,580447

Tabela 2. PSNR das imagens ampliadas 4x

	Vizinho mais próximo	Bilinear	Bicúbico	Zoom com Rede Neural
Imagem a	14,571798	14,296778	15,423591	13,085989
Imagem b	23,746183	23,530838	25,864175	19,812576
Imagem c	24,294603	24,069869	25,373772	21,813326

A seguir, a Figura 9 estabelece uma comparação visual entre a diferentes técnicas utilizadas no processo de ampliação.



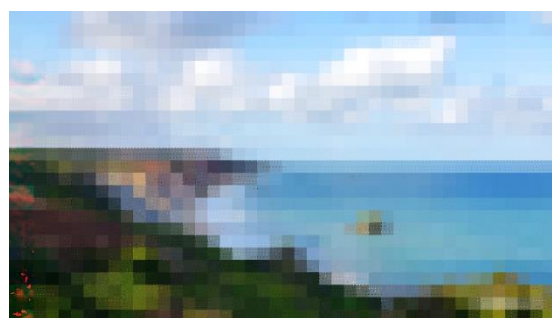
Vizinho mais próximo



Bilinear



Bicúbico



RNA

Figura 9. Ampliação em 4x utilizando diferentes métodos

Fonte: Autoria própria

É perceptível que o método proposto não apresentou resultados satisfatórios com a RNA utilizada (Figura 5). Na Figura 9 o resultado apresenta perda de definição e aspecto serrilhado, deteriorando a qualidade da imagem mesmo em relação ao método do vizinho mais próximo, o mais simples dentre os abordados. Os valores de PSNR apresentados nas Tabelas 1 e 2 reforçam as impressões visuais descritas.

Os testes apresentados foram realizados em um *Virtual Private Server (VPS)* na plataforma *Microsoft Azure*. A máquina virtual dispunha de 4 processadores Intel® Xeon® Platinum 8168 e 8GB de memória *RAM*. Devido ao processo de treinamento da rede neural realizado múltiplas vezes em cada imagem, o método apresentou alta complexidade computacional, tornando sua execução inviável em computadores pessoais. Mesmo utilizando o *VPS*, a execução do método pode durar horas, dias ou semanas, conforme a resolução da imagem e o tamanho da rede neural artificial.

5. Considerações Finais

O presente estudo propôs um método para ampliação de imagens baseado em rede neural artificial multicamadas, posicionando a rede neural como uma alternativa às técnicas de interpolação já estabelecidas na área. O uso desta técnica de inteligência artificial abre uma vasta gama de possibilidades no tratamento do problema, uma vez que o desempenho das redes neurais depende de diversos fatores como a quantidade de neurônios artificiais, número de camadas, funções de ativação, algoritmo de treinamento, iterações no treinamento, dentre outros. Logo, é possível obter diferentes resultados com o método apenas realizando ajustes e/ou modificações da rede neural.

No entanto, o uso de redes neurais eleva bastante a complexidade computacional do método. Em cada imagem, a rede neural é treinada diversas vezes, variando conforme o tamanho da imagem processada. Em cada etapa do treinamento, centenas ou até milhares de neurônios têm que ser processados milhares de vezes. Quanto mais passos são realizados durante o treinamento, melhor a taxa de acerto da rede; quanto maior a quantidade de neurônios que compõem a rede, ela se torna capaz de tratar problemas mais complexos.

A Figura 7 ilustra bem essa situação, onde quanto maior a rede neural utilizada no processamento da mesma imagem, melhor a qualidade do resultado. O processamento de imagens digitais também demanda certo esforço computacional, já que uma imagem pode ser composta por milhões de *pixels*. Cada *pixel*, por sua vez, possui três componentes de cores (RGB) e cada componente de cor deve ser tratado individualmente.

Os resultados apresentados na Figura 9 e nas Tabelas 1 e 2 mostram que a qualidade obtida não é satisfatória e as imagens foram, na verdade, deterioradas pelo método, pois apresentam grande intensidade de ruído. A arquitetura utilizada na rede neural (Figura 5) não se mostrou adequada para tratar o problema, mas os resultados da Figura 7 indicam potencial de sucesso utilizando redes neurais maiores.

Durante o desenvolvimento deste estudo foram encontradas dificuldades para realizar o processamento do método desenvolvido. A complexidade computacional do algoritmo se tornou muito elevada e a indisponibilidade de recursos adequados para sua execução limitou a realização de testes, pois o uso de redes neurais mais complexas não produziria resultados em tempo hábil. Foi utilizado um *Virtual Private Server* na plataforma *Microsoft Azure*, o que auxiliou a execução dos testes, mas ainda não dispunha de recursos apropriados para execução de redes neurais.

Futuramente, pretende-se realizar mais testes com diferentes configurações de redes neurais executadas em computadores adequados. Mesmo com longo tempo de

processamento, o método proposto pode ser potencialmente aplicado à reconstrução de imagens ou outras aplicações que não exijam execução em tempo real.

Referências

- Gonzalez, R. C. e Woods, R. E. (2010). “Processamento Digital de Imagens”. 3ª ed. São Paulo.
- Haykin, S. (2001). “Redes Neurais: Princípios e Prática”. 2ª Ed. São Paulo.
- Hasan, M. S. e Haque, S. T. (2017). “Single Image Super-Resolution Using Back-propagation Neural Network”. Disponível: <https://ieeexplore.ieee.org/document/82818/authors>. Acesso em: 03 de junho.
- Pandey, R. K., Maiya, S. R., e Ramakrishnan, A. G. (2017). “A New Approach for Upscaling Document Images for Improving Their Quality”. Disponível em: <https://ieeexplore.ieee.org/document/8487796>. Acesso em: 03 de junho.
- Russell, S. e Norvig, P. (2010). “Inteligência Artificial”. 3ª ed. Rio de Janeiro.
- Ye, M. , Lyu, D., e Chen, G. Rundo, F. (2020). “Scale-Iterative Upscaling Network for Image Deblurring”. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8963625>. Acesso em: 03 de junho.
- Zhang, X. C., Chen, Q. Ng, R., and Koltun, V. (2019). “Zoom to Learn, Learn to Zoom”. Disponível em: <https://arxiv.org/abs/1905.05169>. Acesso em: 03 de junho.