Similaridades em Imagens Geradas por Redes Neurais Pix2Pix medidas pela rede Xception

Jader dos S. T. Cordeiro, José H. Saito

Centro Universitário Campo Limpo Paulista (UNIFACCAMP)
13231-230 – Campo Limpo Paulista – SP – Brasil

jaderteles@gmail.com, saito@cc.faccamp.br

Abstract. *GAN structure is composed with a pair (G,D) of neural networks, and is used to synthesize images using G, and to evaluate the results using D. In this work it is used a GAN, Pix2pix, to synthesize digital impressions images. To verify the similarities between the synthesized and original images, it is used a convolutional network Xception. The obtained results with 130 images of 13 persons resulted in a high average similaritiy index of 0.9185, with a standard deviation of 0.2273.*

Resumo. *A estrutura GAN, constituída de duas redes neurais, geradora G e discriminadora D, é usada para sintetizar imagens usando G, e aferindo o resultado usando D. Neste artigo é utilizada a rede GAN Pix2pix para gerar imagens sintéticas de impressões digitais. Para verificar a similaridade das imagens sintetizadas com as originais é utilizada uma rede convolucional Xception. Os resultados obtidos com 130 imagens de 13 pessoas resultaram numa alta similaridade média de 0,9185 com desvio padrão de 0,2273.*

1. Introdução

Com o avanço na área de redes neurais artificiais, tornou-se possível a criação de imagens sintéticas de fontes reais, com qualidade realística, muitas vezes indistinguíveis pela percepção humana e podem ser utilizadas para fins diversos. GANs (*Generative Adversarial Networks*) constituem uma classe de estruturas projetadas inicialmente por Goodfellow et al. (2014), em que duas redes neurais confrontam-se entre si: a rede geradora (G) tenta gerar novas imagens a partir de um vetor aleatório, e uma rede discriminatória (D) verifica se essas imagens são reais ou falsas. Após um certo número de iterações, a rede G é capaz de gerar imagens bem próximas das verdadeiras, a ponto da rede D não conseguir distinguir se as imagens geradas são falsas ou verdadeiras. Neste trabalho, aplicamos GANs para gerar imagens sintéticas de impressões digitais. A pesquisa tem como objetivo contribuir para o aprimoramento do método para a ampliação das amostras de treinamento e testes, quando o banco de dados for insuficiente para essas finalidades.

No restante do texto, na Seção 2, são descritos os trabalhos relacionados ao reconhecimento, detecção e classificação de imagens com variedade de redes GANs. Na sequência, Seção 3, abordamos os métodos propostos; na Seção 4 são descritos os experimentos realizados; e na Seção 5, as conclusões finais.

2. Trabalhos relacionados

Os trabalhos relacionados nesta seção abordam as GANs em soluções específicas para gerar imagens com notável precisão e a maioria dos métodos necessitam de um número elevado de amostras que não estão disponíveis. Esta insuficiência na distribuição de entrada pode ser suprida, ampliando o conjunto de dados que possui quantidades insuficientes. Zhang, L. et al. (2019) abordam o reconhecimento das placas de licença de veículos em condições complexas, fundamental no monitoramento rodoviário. Já Ma et al. (2020) utilizam uma nova estrutura de classificação de imagens de células sanguíneas para o diagnóstico na medicina. Na abordagem de Thuy e Hoang (2020) é feita a extração de dados em imagem histopatológica. O aprimoramento dos parâmetros utilizados por Shankaranarayana et al. (2017) resultou em uma melhor triagem do glaucoma como prevenção na perda da visão. A pesquisa de Khan e Mahmoud (2019) em biometria, considera diferenças em sub-grupos de indivíduos, tais como raça, para síntese de imagens e classificação facial. Chugh e Jain (2019) e Wang, G. et al. (2018) evidenciam melhorias no desempenho de contramedidas, nas digitalizações feitas de materiais não vistos durante o treinamento. Avanços importantes foram apresentados por Tan et al. (2019), para classificadores baseados em redes neurais e na detecção de rosto humano falso. Assim, salientamos o importante papel que as redes GANs exercem, nos mais variados segmentos como: monitoramento, saúde, produção, identificação, etc..

3. Métodos propostos

Neste trabalho de pesquisa, utilizamos duas redes neurais convolucionais: a rede GAN Pix2pix de Isola et al. (2017), com a finalidade de gerar imagens sintéticas com qualidade realística a partir das reais e o Xception (*Extreme Exception*) proposta por Chollet (2017), rede neural convolucional – CNN (*Convolutional Neural Network*), a fim de estimar a proporção de similaridade entre as imagens reais e as geradas.

A rede Pix2pix utiliza o recurso de redes adversárias condicionais, as quais não apenas mapeiam a imagem de entrada para a imagem de saída, mas utilizam uma função de perda para treinar esse mapeamento e classificar se a imagem de saída é real ou falsa. Pix2pix possui um gerador baseado na arquitetura U-Net, uma rede convolucional proposta por Ronneberger, Fischer e Brox (2015) e um discriminador que utiliza uma arquitetura PatchGAN, em que a discriminação ocorre por frações de imagens (*patch*). A figura 1 é uma ilustração da rede Pix2pix em diagrama de blocos. O Gerador (G) aprende um mapeamento do vetor de ruído (z) para gerar imagem, e o Discriminador (D) classifica a imagem sintetizada (y) pelo Gerador (G), utilizando um conjunto de amostras reais (x) e determina a probabilidade de serem reais ou falsas.

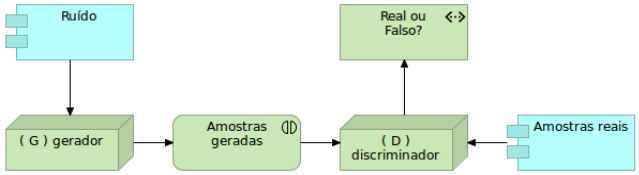

Figura 1. Diagrama de blocos da rede Pix2pix.

A equação (1) da função objetivo mostra que o gerador (G) tenta minimizar a diferença entre a imagem gerada y a partir do vetor de ruído z e a imagem x; e o discriminador (D), maximizar essa diferença.

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x,G(x,z)))] \quad (1)$$

Na equação (2), temos a minimização desse objetivo feita por G, contra a maximização por D, que acontecem em etapas alternadas, com ajustes nos parâmetros na busca de otimização, até um resultado satisfatório.

$$G^* = \arg\min\max \mathcal{L}_{cGAN}(G,D) = \lambda\mathcal{L}_{L1}(G) \quad (2)$$

3.1. Arquiteturas do gerador e discriminador da rede Pix2pix

A rede geradora tem sua estrutura baseada na U-Net de (Ronneberger et al. 2015), sendo inteiramente convolucional e não tem camadas totalmente conectadas. Sua aparência tem um formato de "U" e apresenta uma certa simetria entre as partes. A U-Net é apropriada para associar as informações contextuais que foram obtidas nas camadas de contração, na primeira metade da rede, com os mapas de características e seus equivalentes adquiridos nas camadas de expansão, na segunda metade de rede. As camadas de contração são uma sequência de filtros de convolução 3x3, seguidas da função ReLU (*Rectified Linear Unit*) de (Krizhevsky et al. 2012) e uma operação de max-pooling 2x2 com passo 2. Na camada de expansão, cada etapa consiste na amostragem do mapa de características seguido por uma convolução transposta 2x2, *up-convolution*, uma concatenação com as camadas de contração correspondentes e duas convoluções 3x3, cada uma seguida de uma ReLU. Na última camada, há uma convolução 1x1 que mapeará os vetores de atributos de 64 elementos para o número desejado de classes. A figura 2 mostra a rede geradora, com a correspondência entre as camadas de contração e expansão, de mesmas dimensões, por setas na parte inferior, da arquitetura U-Net. As amostras passam por reduções progressivas, até um limite onde o processo é revertido. Desta forma o fluxo de informações transita por todas as camadas.

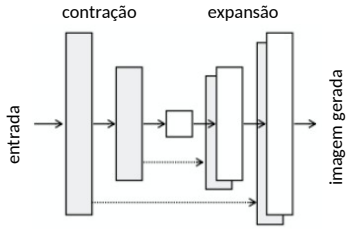

Figura 2. Arquitetura U-Net - adaptado Isola et al. (2017).

O discriminador PatchGAN penaliza a estrutura na escala de frações da imagem (*patches*) determinando se essas frações são reais ou falsas. Para Isola et al. (2017), ao utilizar tamanhos grandes de frações de imagens, o processamento é mais rápido e para otimizar o tempo de processamento é implementado o modelo proposto por Ioffe e Szegedy (2015), um minilote de treinamento (*minibatch*) na normalização de lotes, para conseguir altas taxas de aprendizado.

4. Experimentos

No primeiro experimento (Experimento 1) foi utilizada a rede Pix2pix para a geração de imagens sintéticas a partir de imagens reais. No segundo experimento (Experimento 2) as imagens sintéticas geradas pela rede Pix2pix foram comparadas com as imagens reais utilizando uma rede convolucional Xception proposta por Chollet, F. (2017), para a validação dos resultados. Para o experimento foi utilizado um banco de dados biométrico Sokoto Coventry Fingerprint Dataset Shehu et al. (2018), de impressões digitais, projetado com finalidades acadêmicas, o qual é composto por 6000 imagens de impressões digitais de 600 pessoas. Possui rótulos por gênero, mão esquerda, direita e dedos. Para os experimentos foram utilizadas imagens do dedo indicador da mão direita, na dimensão original de 96x103 pixels. Essas imagens foram ampliadas para 256x256 e criados pares de imagens [z, x] de 512x256 pixels com o software GIMP. Uma das configurações para o treinamento com Pix2pix, é utilizar o formato de pares de imagens [z, x] onde z é uma imagem ruidosa aleatória e x é a imagem de referência. Como recursos computacionais utilizamos a linguagem de programação Python, com TensorFlow, Keras e Ubuntu 20.04, em um sistema computacional i7-9750H com 16GB de RAM e GPU GTX 1660 Ti 6GB.

4.1 Experimento 1

Para o Experimento1, foram montados os conjuntos de dados conforme Tabela 1, onde a primeira coluna identifica os conjuntos P50 a P250, de 5 a 25 pessoas, respectivamente, em que a imagem de cada pessoa é replicada 10 vezes. Após essa replicação, a partir das imagens reais foram obtidas as imagens sintéticas geradas pela rede Pix2pix, para serem medidas as respectivas similaridades no Experimento 2, com a rede Xception. Na figura 3, ilustra-se imagens geradas, pela rede Pix2pix, de impressão digital da Pessoa 05 após 2, 28 e 160 épocas, da esquerda para a direita.

Tabela 1. Conjunto de dados para treinamento

	Treinamento	
	Imagens	Pessoas
P50	50	5
P100	100	10
P150	150	15
P200	200	20
P250	250	25

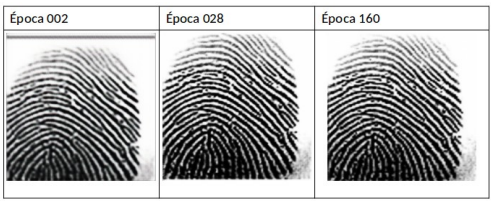

Figura 3. Imagem gerada pela rede Pix2pix da Pessoa 05, após 2, 28 e 160 épocas.

4.2. Experimento 2

Em seguida apresentamos a Tabela 2 com o resultado da aplicação da rede Xception de similaridade, na qual a média e o desvio padrão se atribuem aos resultados da classificação das dez imagens originais e das dez sintetizadas de cada pessoa. Na coluna de resultado das imagens reais, os valores correspondem à media e ao desvio padrão entre as imagens reais respectivas. No entanto, na coluna das imagens sintetizadas, os valores correspondentes são resultantes das comparações entre as imagens reais e as sintetizadas. Em ambos os casos, os valores de média próximos a 1 demonstram maior grau de similaridade.

Tabela 2. Proporção de similaridade dos conjuntos

	Teste		Similaridade das imagens reais		Similaridade das imagens sintetizadas	
Conjunto	Imagens	Pessoas	Média	Desvio padrão	Média	Desvio padrão
P50	30	3	0,9924	0,0060	0,8794	0,0990
P100	50	5	1,0000	0,0000	0,9998	0,0003
P150	80	8	1,0000	0,0000	1,0000	0,0000
P200	100	10	1,0000	0,0000	0,9973	0,0055
P250	130	13	1,0000	0,0000	0,9185	0,2273

Observamos que os resultados foram influenciados pelo aumento dos conjuntos treinados, com réplicas das mesmas imagens. A identificação das impressões digitais por pessoa, teve boa taxa de acerto, com exceção do conjunto P50, em que a média de similaridade foi de 0,8794 para imagens sintetizadas, possivelmente pela quantidade pequena de amostras utilizadas para treinamento. Para os grupos P100, P150 e P200 os resultados de similaridade foram próximos de 1, com desvio padrão próximo de 0. Foi verificado um valor de similaridade 0,9185 com desvio padrão para 0,2273 para o conjunto P250, provavelmente devido ao número grande de pessoas envolvidas.

5. Conclusões finais

A versatilidade da rede GAN tem se apresentado como uma solução promissora nas mais diversas áreas, possibilitando o aumento de dados com imagens sintéticas acrescidas às imagens reais, contribuindo para a ampliação dos bancos de dados. Neste trabalho realizamos a sintetização de imagens de impressões digitais, partindo de um conjunto reduzido de amostras de imagens reais, com a rede GAN Pix2pix. Foram verificadas usando a rede Xception que para impressões digitais com até 10 pessoas envolvidas, os índices de similaridade das imagens sintetizadas com as imagens reais são próximos de 1. Como trabalhos futuros, pretende-se pesquisar outros tipos de redes GAN, além de outros métodos de sintetização de imagens para finalidades de geração de imagens.

Referências

Chollet, F. (2017). "Xception: Deep Learning with Depthwise Separable Convolutions". In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 1800-1807, doi: 10.1109/CVPR.2017.195.

Chugh, T. e Jain, A. K. (2019). Fingerprint Spoof Detector Generalization. In: *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 42-55, 2021, doi: 10.1109/TIFS.2020.2990789.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. e Bengio, Y. (2014). “Generative adversarial nets”. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14). MIT Press, Cambridge, MA, USA, 2672–2680.

Ioffe, S. e Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning* - Volume 37 (ICML'15). JMLR.org, 448–456.

Isola, P., Zhu, J., Zhou, T. e Efros, A. A. (2017). "Image-to-Image Translation with Conditional Adversarial Networks" In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 5967-5976.

Khan, A. e Mahmoud, M. (2019). "Considering Race a Problem of Transfer Learning". In: 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), Waikoloa Village, HI, USA, 2019, pp. 100-106, doi: 10.1109/WACVW.2019.00022.

Krizhevsky, A., Sutskever, I., e Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks". Communications of the ACM 60(6): 84–90.

Ma, L., Shuai, R., Ran, X., Liu, W., Ye, C. (2020). Combining DC-GAN with ResNet for blood cell image classification. In: *Med Biol Eng Comput*, 58**,** 1251–1264 (2020). doi: 10.1007/s11517-020-02163-3.

Ronneberger O., Fischer P. e Brox T. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham.

Shankaranarayana S.M., Ram K., Mitra K. e Sivaprakasam M. (2017). "Joint Optic Disc and Cup Segmentation Using Fully Convolutional and Adversarial Networks". In: Cardoso M. et al. (eds) Fetal, Infant and Ophthalmic Medical Image Analysis. OMIA 2017, FIFI 2017. Lecture Notes in Computer Science, vol 10554. Springer, Cham, doi: 10.1007/978-3-319-67561-9_19.

Shehu, Y., I., Ruiz-Garcia, A., Palade, V. e James, A. (2018) "Sokoto Coventry Fingerprint Dataset". arXiv, vol 1807.10609.

Tan T., Wang X., Fang Y. e Zhang W. (2019). "The Impact of Data Correlation on Identification of Computer-Generated Face Images". In: Sun Z., He R., Feng J., Shan S., Guo Z. (eds) Biometric Recognition. CCBR 2019. Lecture Notes in Computer Science, vol 11818. Springer, Cham, doi:10.1007/978-3-030-31456-9_17.

Thuy M.B.H. e Hoang V.T. (2020). "Fusing of Deep Learning, Transfer Learning and GAN for Breast Cancer Histopathological Image Classification". In: Le Thi H., Le H., Pham Dinh T., Nguyen N. (eds) Advanced Computational Methods for Knowledge Engineering. ICCSAMA 2019. Advances in Intelligent Systems and Computing, vol 1121. Springer, Cham, doi: 10.1007/978-3-030-38364-0_23.

Wang, G., Kang, W., Wu, Q., Wang, Z. e Gao, J. (2018). Generative Adversarial Network (GAN) Based Data Augmentation for Palmprint Recognition. In: *2018 Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, Australia, 2018, pp. 1-7, doi:10.1109/DICTA.2018.8615782.

Zhang L., Wang P., Dang F. e Zhang S. (2019). "A Simple and Robust Attentional Encoder-Decoder Model for License Plate Recognition". In: Lin Z. et al. (eds) Pattern Recognition and Computer Vision. PRCV 2019. Lecture Notes in Computer Science, vol 11857. Springer, Cham, doi: 10.1007/978-3-030-31654-9_26.