

Análise de redes neurais adversariais generativas para a geração de imagens sintéticas

Marcus V. D. Zuba
 Rogério M. Gomes¹
 Bruno A. Santos
 DECOM-CEFET-MG, Belo Horizonte, MG

As Generative Adversarial Networks (GANs) ou redes neurais adversariais generativas, são modelos de redes neurais artificiais que têm sido usadas na geração de instâncias sintéticas de dados, incluindo imagens, vídeos, textos, entre outras. As GANs são sistemas compostos por dois modelos de redes neurais concorrentes que competem entre si e são capazes de analisar, capturar e copiar as variações em um conjunto de dados, gerando dados sintéticos que são semelhantes aos dados reais de treinamento, como ilustrado na Figura 1 [1].

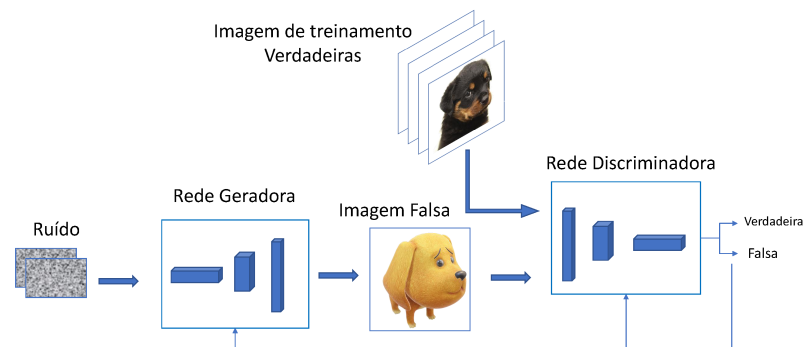


Figura 1: Estrutura geral das redes GANs.

A função da rede generativa é produzir imagens que se pareçam ao máximo com as reais, enquanto a função da rede discriminativa é julgar se as imagens são verdadeiras ou falsas. O treinamento das duas redes é realizado em conjunto, ou seja, correções pontuais são feitas, tanto na rede discriminadora quanto na geradora, tendo como base o resultados obtidos pelo discriminador.

As GANs originais possuem a rede geradora e discriminadora constituídas por camadas lineares que são limitadas, principalmente quando as imagens são coloridas. A DCGAN (Deep Convolutional Generative Adversarial Networks) [2] substituiu as camadas lineares das GANs por camadas convolucionais tradicionais na rede discriminadora e camadas convolucionais transpostas na rede geradora, melhorando o FID das imagens geradas. A métrica FID (Frechet Inception Distance), usada para avaliar o desempenho das redes, é responsável por calcular a distância entre as imagens geradas e as imagens reais de um mesmo domínio.

A rede geradora do modelo BigGAN [3], por sua vez, além do vetor de ruídos usado na geração da imagem, utiliza também um vetor das classes de objetos que serão gerados. O modelo também

¹rogerio@cefetmg.br.

possui uma rede neural auxiliar Self-Attention, responsável por obter características de diferentes partes das imagens a serem utilizadas, tanto na rede geradora quanto na discriminadora. A BigGAN, visando melhorar a estabilidade da rede e a geração de imagens sintéticas de alta qualidade, implementou diversas outras melhorias, a saber: Normalização Espectral; Atualização do discriminador mais do que o gerador; Regularização ortogonal; entre outras.

Este projeto tem por objetivo avaliar o impacto da configuração dos hiperparâmetros no desempenho dos modelos DCGAN e BigGAN aplicados a base de dados disponibilizada pelo sítio Kaggle na competição *Generative Dog Images*², que possui 20.580 imagens coloridas, com uma resolução 64x64 pixels, de 120 raças de cachorros.

Os resultados dos experimentos realizados com a melhor seleção de hiperparâmetros podem ser vistos na Figura 2. É possível observar que a rede BigGAN apresentou um melhor desempenho (menor valor de FID) em comparação com as redes DCGAN. Sendo assim, podemos concluir que apesar da rede BigGAN ter incluído uma maior complexidade às redes GANs, ela apresentou, além de uma melhor estabilidade, um alto desempenho na geração de imagens sintéticas.



Figura 2: Resultados visuais e de FID obtidos pelas redes (a) DCGAN e (b) BigGAN

Agradecimentos

Os autores gostariam de agradecer ao CEFET-MG pelo suporte financeiro.

Referências

- [1] Goodfellow, I. J., Pouget-Abadie, J., Mirza M., Xu, B., Warde-Farley D., Ozair S., Courville A., and Bengio Y. Generative adversarial nets. *stat*, 1050:10, 2014.
- [2] Radford A., Metz L., and Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [3] Brock, A., Donahue, J. and Simonyan, K. Large scale gan training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2018.

²<https://www.kaggle.com/c/generative-dog-images> - Acessado em 27/04/2021)