

Análise de Lesões de Pele usando Redes Generativas Adversariais

Alceu Bissoto¹, Sandra Avila¹

¹Instituto de Computação, Universidade Estadual de Campinas (UNICAMP), Brasil

{alceubissoto, sandra}@ic.unicamp.br

Resumo. *O câncer de pele é de longe o tipo mais comum de câncer. O diagnóstico precoce é fundamental para o tratamento do paciente, melhorando significativamente as taxas de sobrevivência. O aprendizado profundo tornou-se o estado da arte na análise de lesões de pele, e os dados se tornaram um fator decisivo para impulsionar as soluções. O objetivo principal desta dissertação de mestrado é abordar dois problemas que surgem ao ter conjuntos de dados limitados: a incapacidade dos modelos de generalizar e o risco de viés do conjunto de dados. Investigamos os dois problemas separadamente, propondo métodos inovadores que contribuíram para o avanço da análise de lesões de pele.*

1. Introdução

Melanoma é o mais perigoso tipo de câncer de pele. Ele causa o maior número de mortes, mesmo representando por volta de apenas 1% de todos os cânceres de pele nos Estados Unidos¹, e 3% no Brasil². O ponto crucial para o tratamento do melanoma é a detecção precoce. A taxa de sobrevivência estimada em 5 anos dos pacientes diagnosticados aumenta de 15%, se detectado em seu estágio mais avançado, para mais de 97% se detectado em seus estágios iniciais.

O procedimento de diagnóstico médico para melanoma baseia-se na análise de padrões, permitindo que o aprendizado de máquina seja uma abordagem prática para esse problema. As redes neurais profundas são o estado da arte para análise de lesões de pele, e dados são críticos para melhorar essas soluções. Para contextos médicos, como o nosso, a falta de dados anotados é um problema grave. Adquirir e rotular novas amostras é caro (em dinheiro e tempo), fazendo conjuntos de dados disponíveis serem muito limitados.

Essa severa limitação em nossos dados cria dois problemas cruciais investigados neste mestrado: a incapacidade de generalização das soluções, e o viés do conjunto de dados. Para lidar com o primeiro problema, apresentamos um método baseado em Redes Generativas Adversariais (*Generative Adversarial Networks*, GANs) para gerar dados sintéticos realistas com o intuito de melhorar a generalização dos modelos de classificação de lesões de pele. Nosso método foi capaz de gerar imagens rotuladas de alta resolução, clinicamente relevantes, que quando usadas para compor os conjuntos de dados de treinamento para classificação, aumentaram significativamente o desempenho. Para o segundo problema, projetamos experimentos para estudar como nossos modelos de classificação usam os dados, verificando como eles exploram correlações corretas (médicas) e espúrias. Perturbadoramente, mesmo na ausência de informações clínicas sobre a lesão a ser diagnosticada, nossos modelos de classificação tiveram um desempenho preciso, sugerindo um desempenho inflado dos modelos pelo viés no conjunto de dados.

¹<http://www.cancer.net/cancer-types/melanoma/statistics>

²<https://www.inca.gov.br/tipos-de-cancer/cancer-de-pele-melanoma>

2. Dados

Existem dois tipos de imagens de lesões de pele: dermatoscópicas e clínicas. As imagens clínicas podem ser capturadas com câmeras comuns, enquanto as imagens dermatoscópicas só podem ser capturadas com um dispositivo chamado dermatoscópio. Este dispositivo normaliza a influência da luz na lesão, permitindo capturar detalhes mais profundos. As imagens dermatoscópicas permitem a aplicação de algoritmos médicos que auxiliam a decisão dos especialistas. Alguns destes algoritmos se baseiam em características da lesão, como a regra ABCD (assimetria, borda, cor e diâmetro das lesões); e outros, como a regra de 7 pontos, se baseiam em **atributos dermatoscópicos**. Esses atributos estão presentes nas lesões na forma de padrões visuais, como redes, glóbulos e estrias. Há uma grande variedade de atributos dermatoscópicos e cada um deles pode estratificar de acordo com a sua regularidade, cor e outras especificidades.

Ao longo deste mestrado, contamos com atributos dermatoscópicos para gerar imagens de lesões de pele e medir o viés em nossos conjuntos de dados de lesões de pele.

3. Síntese de Imagens de Lesões de Pele

As GANs visam modelar a distribuição real dos dados, forçando as amostras sintéticas a serem indistinguíveis dos dados reais. Neste mestrado propusemos um método baseado em GANs para gerar imagens sintéticas de lesões de pele em alta definição, visualmente coerentes e clinicamente significativas. Este trabalho foi o primeiro a gerar com sucesso imagens realistas de lesões de pele (Fig. 1). Para avaliar a relevância das imagens sintéticas, treinamos uma rede de classificação de câncer de pele com imagens sintéticas e reais, alcançando uma melhora de 1.3 ponto percentual. Nosso código está disponível em <https://github.com/alceubissoto/gan-skin-lesion>.

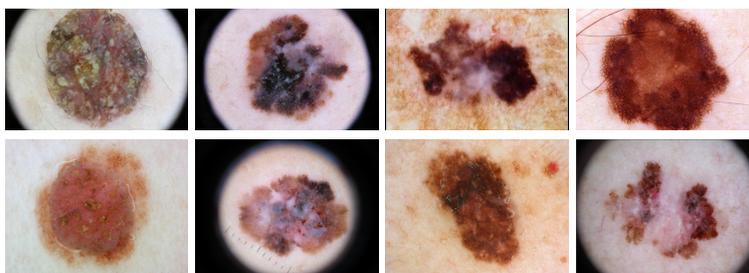


Figura 1. Comparação entre nossas amostras sintéticas (linha superior) e amostras reais da base de dados *ISIC Archive* (linha inferior).

Nosso objetivo é gerar imagens sintéticas de alta resolução de lesões de pele com detalhes refinados que possam fornecer correlações corretas às redes de classificação, auxiliando em sua generalização. Para isso, alimentamos diretamente o gerador da GAN com mapas que codificam atributos dermoscópicos das lesões e as especificidades das bordas. Dessa forma, em vez de gerar a imagem a partir de ruído (procedimento usual com GANs), sintetizamos a partir de um mapa de rótulos semânticos e de um mapa de instâncias (Fig. 2). Assim, nosso problema de síntese de imagem pode ser representado como um problema de tradução de imagem em imagem. A Fig. 3 resume nosso *pipeline*.

Os mapas semânticos e de instância são cruciais. Eles encapsulam informações sobre as lesões de pele que simplificam o processo de geração, permitindo a síntese de

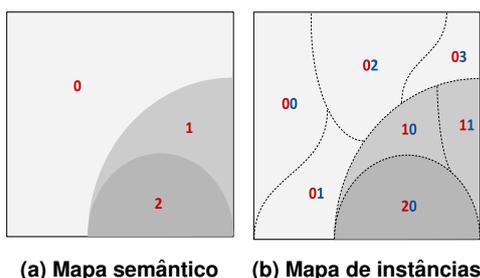


Figura 2. Simplificação de nossos mapas semânticos e de instância. Enquanto os valores dos pixels do mapa semântico são regidos apenas pela classe, os mapas de instância levam em consideração a classe e a instância individual definida por *superpixels*.

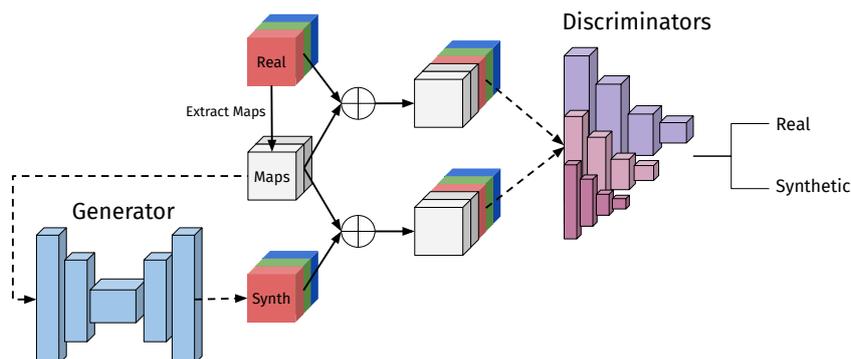


Figura 3. Nosso *pipeline*. Alimentamos o gerador com mapas extraídos de imagens reais, resultando em imagens sintéticas. O discriminador é alimentado com *batches* que combinam imagens reais e seus mapas, ou imagens sintéticas e os mapas usados para gerá-los. A saída dos discriminadores (existem três, cada um operando em uma resolução diferente) é finalmente retropropagada para treinar todo o *pipeline*.

imagens de maior qualidade. Ambos os mapas são baseados em máscaras de segmentação e anotações de atributos dermatoscópicos fornecidas publicamente por especialistas como parte do *ISIC Challenge 2018*. A anotação compreende 5 tipos de padrões dermatoscópicos (rede de pigmentos, rede negativa, estrias, cistos semelhantes a milia, e glóbulos), fornecendo uma máscara binária para cada um.

Para avaliar o conjunto completo de imagens sintéticas, treinamos uma rede de classificação de pele com diferentes combinações de dados sintéticos e reais (incluindo apenas reais, apenas sintéticos, e combinações de ambos) para compor nosso conjunto de dados de treinamento. Comparamos a área alcançada sob a curva ROC (AUC), testando sempre apenas com imagens reais. Os resultados para o treinamento da rede de classificação com imagens sintéticas confirmam que eles contêm características que diferenciam uma lesão como maligna ou benigna, e sua adição ao conjunto de treinamento de uma rede de classificação melhora o desempenho em AUC a uma média de 1,3% pontos percentuais, além de manter a rede mais estável.

Destacamos que esta contribuição foi publicada [Bissoto et al. 2018c], e ainda é, até hoje, o estado da arte para a síntese de imagens de lesões de pele.

4. Vieses em Conjuntos de Dados de Lesões de Pele

Devido à escassez de imagens de lesões de pele anotadas e de boa qualidade, dois conjuntos de dados dominam a pesquisa em análise automática de lesões de pele: o *Interactive*

Atlas of Dermoscopy e o *ISIC Archive*. Os vieses dos conjuntos de dados podem aumentar o desempenho dos modelos (apresentando características que não são verdadeiros aos dados do mundo real) ou diminuir o desempenho deles (destruindo correlações coerentes com o mundo real, e impedindo que os modelos as explorem).

Apesar de ser impossível de eliminar completamente, é importante entender o viés e suas fontes para melhorar nossos processos de aquisição de imagens e modelos de aprendizado profundo. Uma maneira útil de medir o primeiro efeito possível de viés em um conjunto de dados (desempenhos indevidamente inflados por conta das correlações espúrias no conjunto de dados) é um experimento contrafactual, que destrói as informações importantes nos dados e mede quanto o desempenho dos modelos diminui. Portanto, nosso primeiro conjunto de experimentos segue esse procedimento, removendo gradualmente as informações das imagens das lesões de pele e medindo o desempenho da rede. Realizamos experimentos em conjuntos de dados únicos (treinamento e teste no mesmo conjunto de dados) e cruzados (treinamento no *ISIC* e teste no *Atlas*) e descobrimos que em ambos os casos, as redes são capazes de manter um surpreendente desempenho, mesmo depois de quase todas as informações relevantes serem destruídas (Fig. 4).

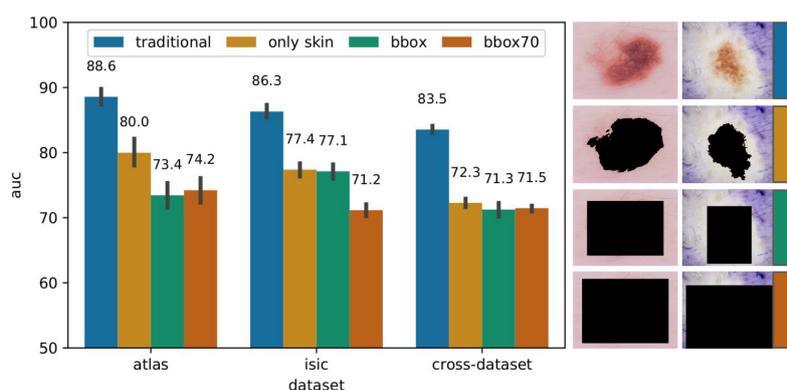


Figura 4. Desempenho dos modelos nos conjuntos de dados alterados. Primeiro, removemos todas as cores de dentro da lesão (*only skin*), depois as informações da borda (*bbox*) e, finalmente, removemos o tamanho (diâmetro) da lesão (*bbox70*). Mostramos amostras de cada conjunto de dados à direita. Surpreendentemente, mesmo quando destruímos todas as informações clínicas relevantes, a rede encontra uma maneira de aprender a classificar as imagens de lesões de pele muito melhor que o acaso.

Medir o segundo efeito possível (incapacidade de fornecer correlações úteis para a aprendizagem) é muito mais difícil, pois não podemos provar, *a priori*, que essas correlações existem no mundo real, nem que o modelo de aprendizado de máquina aprenderia com elas se eles foram representados corretamente no conjunto de dados. O melhor que podemos fazer é fornecer evidências adicionais para os modelos que esperamos que sejam úteis para um humano e avaliar se isso faz alguma diferença.

Assim, em nosso segundo conjunto de experimentos, adicionamos progressivamente mais informações, com base em atributos dermatoscópicos refinados (rede de pigmentos, rede negativa, estrias, cistos semelhantes a milia, e glóbulos) em forma de mapas. Para fornecer esses recursos, empregamos as anotações disponíveis do *ISIC Challenge 2018*. Esperávamos que essas clinicamente relevantes informações de lesões de pele melhorassem o processo de aprendizado da rede, mas, na verdade, o desempenho não me-

hora em todos os cenários testados, mesmo quando alimentamos a rede com todos os pixels da imagem, mais um canal adicional contendo informações clínicas (Fig. 5).

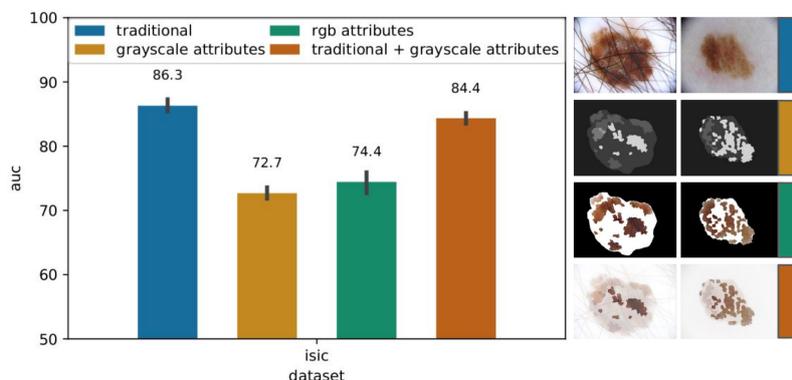


Figura 5. Comparação de desempenho dos diferentes conjuntos de imagens utilizando o conjunto de dados *ISIC*. Mostramos amostras de cada conjunto à direita. Surpreendentemente, quando tentamos simplificar o processo de aprendizado alimentando a rede com atributos dermatoscópicos, o resultado não melhora.

Quando ocultamos informações relevantes das lesões dos nossos modelos, eles ainda deveriam aprender padrões que ajudam a diferenciar lesões benignas de malignas? Acreditamos que quando um modelo aprende a classificar lesões malignas analisando apenas a pele — sem informações nas bordas, marcadores biológicos ou diâmetro das lesões —, ele se apoia fortemente em artefatos introduzidos durante a aquisição da imagem e no viés geral do conjunto de dados. Esse problema é crítico para implantar a análise automática de lesões de pele. Ao atuar no mundo real, queremos que a rede seja o mais imparcial possível, tomando decisões com base em recursos clínicos. Portanto, é urgente entender o viés nos conjuntos de dados usados para treinar e avaliar nossos modelos.

Destacamos que essa contribuição foi publicada [Bissoto et al. 2019a], e o nosso artigo recebeu o prêmio *Best Paper Award*. O nosso código está disponível em <https://github.com/alceubissoto/deconstructing-bias-skin-lesion>.

5. Resultados da Pesquisa

- Dois artigos de conferência [Bissoto et al. 2018c, Bissoto et al. 2019a] e um relatório técnico [Bissoto et al. 2019b].
- Premiação de segundo melhor pôster no *International Educational Symposium of The Melanoma World Society* [Bissoto et al. 2018a].
- Sexto lugar no *ISIC 2018 Challenge* [Bissoto et al. 2018b].
- *Best paper award* no *ISIC Skin Image Analysis Workshop*, na *Conference on Computer Vision and Pattern Recognition (CVPR)* [Bissoto et al. 2019a].
- Vencedor do *Google Latin America Research Awards* por dois anos consecutivos (Google LARA 2018 and 2019).
- Disseminação dos resultados da nossa pesquisa na mídia tradicional:
 - Jornal Estadão <https://tinyurl.com/v8fzqnl>
 - Jornal Correio Popular <https://tinyurl.com/sepv732>
 - Jornal da Unicamp <https://tinyurl.com/s76ggdo>
 - TV Cultura <https://tinyurl.com/vc7qbad>

- EPTV 2ª Edição <https://tinyurl.com/tov4vj4>
- TVB Record TV Campinas <https://tinyurl.com/rpabxod>
- TV BandMais <https://tinyurl.com/u7cwr2w>
- Rádio CBN, Show da Notícia <https://tinyurl.com/tve42ff>

6. Conclusões e Trabalhos Futuros

- **Revisão da literatura de GAN:** Fornecemos um estado da arte abrangente de GANs, dividindo-a em seis tópicos, mostrando como os trabalhos se influenciaram até chegarmos ao estágio em que estamos hoje [Bissoto et al. 2019b].
- **Síntese da imagem de lesões de pele:** Propusemos um método para a síntese de lesões de pele que gera lesões de pele sintéticas clinicamente relevantes de alta definição. Nosso método é, até hoje, o estado da arte em síntese de lesões de pele, e temos muitos caminhos para aumentar a qualidade e a variabilidade de nossas amostras sintéticas [Bissoto et al. 2018a, Bissoto et al. 2018c].
- **Viés nos conjuntos de dados de lesões de pele:** Investigamos os dados usados para os modelos de síntese e classificação. O desempenho do modelo de classificação, mesmo quando nenhuma informação clinicamente significativa é apresentada (de acordo com algoritmos médicos), é surpreendentemente alta, superando *benchmarks* que quantificam o desempenho de especialistas [Bissoto et al. 2019a]. Nosso trabalho aumentou a conscientização da comunidade quanto aos vieses, e esperamos que isto possa ser solucionado em breve.

O câncer já é um desafio de saúde pública. Prevê-se que 30 milhões de pessoas morrerão de câncer por ano até 2030. A grande maioria das mortes ocorrerá em países de baixa-média renda, como o Brasil. Soluções automatizadas, como a explorada nesta pesquisa, podem ajudar a mudar esse horizonte, alcançando e selecionando as pessoas que estão em risco e permitindo que elas sejam o foco de especialistas.

Acreditamos que podemos acelerar esse processo extraindo o máximo de cada amostra e combinando diferentes domínios (texto de registros médicos, imagens clínicas e histopatológicas, e genômica). Reunir dados de qualidade diversificados e aprender a combinar e entender esses dados pode ser a próxima revolução na análise automatizada.

Referências

- Bissoto, A., Fornaciali, M., Valle, E., and Avila, S. (2018a). Generating high quality synthetic skin lesions for boosting automated screening. *International Educational Symposium of the Melanoma World Society*, pages 43–43.
- Bissoto, A., Fornaciali, M., Valle, E., and Avila, S. (2019a). (De)Constructing bias on skin lesion datasets. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*.
- Bissoto, A., Perez, F., Ribeiro, V., Fornaciali, M., Avila, S., and Valle, E. (2018b). Deep-learning ensembles for skin-lesion segmentation, analysis, classification: RECOD Titans at ISIC Challenge 2018. *arXiv preprint arXiv:1808.08480*.
- Bissoto, A., Perez, F., Valle, E., and Avila, S. (2018c). Skin lesion synthesis with generative adversarial networks. In *ISIC Skin Image Analysis Workshop, International Conference on Medical Image Computing and Computer Assisted Intervention*.
- Bissoto, A., Valle, E., and Avila, S. (2019b). The six fronts of the generative adversarial networks. *arXiv preprint arXiv:1910.13076*.