

Análise de funções de perda para segmentação de lesões cutâneas usando redes neurais convolucionais e transformações polares

E. D. Bitencourt¹, J. K. S. Kamassury¹ e D. Silva¹

¹ Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina, Florianópolis, Brasil

Abstract— Automated skin lesion segmentation in dermoscopy images is important for disease diagnosis and treatment planning. The use of convolutional neural networks for this purpose, especially using recently proposed polar transformations, has shown high potential and performance. This paper generalizes the method for different loss functions, showing that it is relatively insensitive to this choice, with region-based losses being slightly superior to binary cross-entropy in terms of the Dice coefficient.

Keywords— Polar transformations, Deep learning, Loss functions, Semantic segmentation, ISIC.

I. INTRODUÇÃO

Redes neurais convolucionais [1] têm proporcionado significativos progressos na predição de diagnósticos médicos, especialmente na área de segmentação de lesões cutâneas. Nesse panorama, as edições do desafio anual *International Skin Imaging Collaboration* (ISIC) têm contribuído notavelmente para a continuidade desse progresso.

Com relação à tarefa de segmentação de conjunto de dados de 2018 [2], o método atual mais bem-sucedido consiste em alterar os eixos das imagens via transformações polares [3]. A intuição desse método reside em considerar que as imagens das lesões possuem formatos com distribuições aproximadamente elípticas, o que permite simplificar o formato da segmentação e alavancar o processo de aprendizado.

Apesar de terem obtido valores elevados nas métricas *Dice similarity coefficient* (DSC) e *median intersection-over-union* (mIoU), de 0,9224 e 0,8699 respectivamente, observa-se que somente a função de perda *Dice Loss* (DL) [4] foi usada nessa abordagem. Isto posto, o presente trabalho dedica-se em obter e comparar os desempenhos desta abordagem quando outras funções de perda (incluindo funções robustas ao desbalançamento de classes) são empregadas para o treinamento do modelo de segmentação.

II. ESTADO DA ARTE

Para a realização da transformação polar, o maior desafio é definir o melhor pixel central da imagem cartesiana para

realizá-la. Dentre as diferentes formas propostas pelo estado da arte [3] para tal propósito, focamos na metodologia em que um modelo treinado estima o ponto central da imagem que, por sua vez, é submetida à transformação polar. Na etapa seguinte, a segmentação (foco deste estudo) é feita por outro modelo que retorna uma predição polar. A Fig. 1 ilustra as etapas desta metodologia.

III. FUNÇÕES DE PERDA PARA SEGMENTAÇÃO

No contexto do aprendizado de máquina, a escolha da função de perda é significativamente importante ao projetar e treinar redes neurais profundas visto que ela define como os modelos calculam o erro geral de seus resíduos para cada lote de treinamento que, por sua vez, afeta como os pesos do modelos são ajustados ao realizarem a retropropagação. Em geral, para tarefas de segmentação semântica, a escolha padrão é a *binary cross-entropy* (BCE) [4] expressa por

$$\mathcal{L}_{bce}(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}) \quad (1)$$

em que $y \in \{0, 1\}$ é a máscara real e \hat{y} é a máscara prevista pela rede neural. Entretanto, sabe-se que devido a BCE ponderar igualmente os *pixels* independentemente da classe, para cenários em que ocorre o desequilíbrio de classes, ela tende a favorecer a classe majoritária e classificar erroneamente a classe-alvo. Dentre as diferentes funções alternativas à BCE, sobressaem aquelas baseadas em região, as quais visam minimizar a incompatibilidade ou maximizar as regiões de sobreposição entre y e \hat{y} . Nesse contexto, citam-se a *jaccard loss* (JL) [5], a DL e a *tversky loss* (TL) [4] que são expressas respectivamente como

$$\mathcal{L}_{jaccard}(y, \hat{y}) = 1 - \frac{y\hat{y} + \epsilon}{y + \hat{y} - y\hat{y} + \epsilon} \quad (2)$$

$$\mathcal{L}_{dice}(y, \hat{y}) = 1 - \frac{2y\hat{y} + \epsilon}{y + \hat{y} + \epsilon} \quad (3)$$

$$\mathcal{L}_{tversky}(y, \hat{y}) = 1 - \frac{y\hat{y} + \epsilon}{y\hat{y} + \alpha(1 - y)\hat{y} + \beta y(1 - \hat{y}) + \epsilon} \quad (4)$$

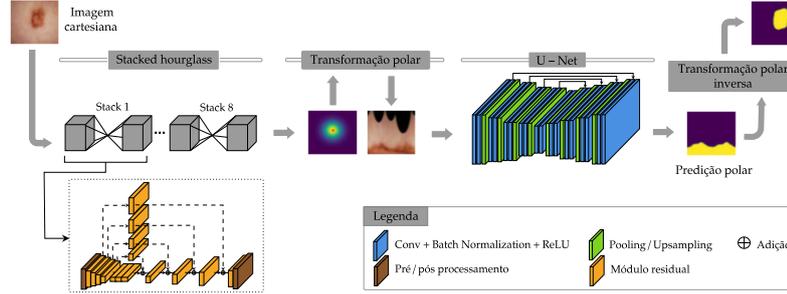


Figura 1: Diagrama da abordagem na qual se usa um modelo preditor de ponto central (*stacked hourglass*) e uma rede de segmentação (U-Net) treinada em imagens oriundas de transformações polares (consulte [3] para detalhes das arquiteturas e transformações polares).

onde ε é usado para garantir estabilidade numérica e os parâmetros α e β ponderam os falsos positivos e falsos negativos. Ainda é válido mencionar a *focal tversky loss* (FTL) [4], que é uma generalização da TL, sendo definida por

$$\mathcal{L}_{focal\ tversky}(y, \hat{y}) = \mathcal{L}_{tversky}(y, \hat{y})^\gamma \quad (5)$$

na qual o parâmetro γ controla a não-linearidade da perda.

IV. EXPERIMENTOS E RESULTADOS

Para investigarmos os desempenhos alcançados com diferentes funções de perda para o modelo de segmentação, inicialmente reproduzimos os resultados do estado da arte [3] para o conjunto de dados ISIC 2018, usando a metodologia discutida na seção II. Em detalhe, esse conjunto de dados (combinando os dados de treinamento e validação) possui 2694 imagens dermatoscópicas com suas respectivas máscaras de segmentação binária rotuladas por especialistas.

Seguindo a abordagem recomendada, divide-se esse conjunto na proporção de 8 : 1 : 1 para o conjunto de treinamento, validação e teste, respectivamente. Além disso, cada imagem é redimensionada para uma resolução de 384×512 pixels e normalizada para um intervalo de $[-0, 5, 0, 5]$.

Na etapa de treinamento, o modelo de segmentação foi treinado por 200 épocas utilizando um *batch size* de 8 e o otimizador Adam [1] com uma taxa de aprendizado de 10^{-3} . Na fase de teste, avaliaram-se os desempenhos do modelo segundo às métricas DSC e mIoU, *precision* e *recall* cujos resultados estão dispostos na Tabela 1. Para as funções de perda que dependem dos hiperparâmetros α , β e γ , usamos os valores 0,3, 0,7 e 3/4, respectivamente.

Os resultados evidenciam que, em comparação com a função DL, é possível obter uma pequena melhora nas métricas Dice e mIoU pelo uso da função TL, a qual concomitantemente demonstra resultados parcialmente equilibrados de *precision* e *recall*.

Tabela 1: Desempenhos do modelo de segmentação treinado com diferentes funções de perda

Funções de perda	Métricas			
	Dice	mIoU	Precision	Recall
DL	92,6	87,6	93,6	93,4
JL	92,7	87,7	93,2	94,1
TL	92,8	87,9	93,9	93,4
FTL	92,7	87,7	94,6	92,4
BCE	92,3	87,0	93,3	92,9

V. CONSIDERAÇÕES FINAIS

Esse artigo generaliza o estado da arte [3] em segmentação de lesões cutâneas e explicita a eficiência das transformações polares para diversas funções de perda dentro do escopo delineado, evidenciando que novos métodos baseados no formato da imagem se mostram com alto potencial. Em particular, o método se mostra relativamente insensível à escolha da função de perda, com funções baseadas em região apresentando um desempenho ligeiramente superior em relação à perda BCE.

REFERÊNCIAS

- Goodfellow Ian, Bengio Yoshua, Courville Aaron. *Deep Learning*. MIT Press 2016.
- Codella Noel, Rotemberg Veronica, Tschandl Philipp, et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC) Tech. Rep. arXiv:1902.03368arXiv 2019.
- Benčević Marin, Galić Irena, Habijan Marija, Babin Danilo. Training on Polar Image Transformations Improves Biomedical Image Segmentation *IEEE Access*. 2021;9:133365-133375.
- Jadon Shruti. A survey of loss functions for semantic segmentation in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*:1–7IEEE 2020.
- Eelbode Tom, Bertels Jeroen, Berman Maxim, et al. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index *IEEE Transactions on Medical Imaging*. 2020;39:3679-3690.