

Segmentação de imagens de Células Blásticas e Não-Blásticas

Felipe Augusto Dittert Noletto^{#1}, Victor Ribeiro Garcia^{*2}

Processamento de Imagens Biomédicas, Universidade Federal do Paraná

¹GRR: 20205689

²GRR: 20203954

Abstract— Este estudo investiga o desempenho do clustering k-means em imagens de células blásticas e não-blásticas usando a base de dados ALL-IDB2. Células blásticas, indicativas de leucemia, e células não-blásticas de indivíduos saudáveis foram avaliadas separadamente. As imagens foram processadas nos espaços de cor RGB e HSV. Os melhores parâmetros de clustering, incluindo o modelo de cor e o número de clusters (k), foram determinados. Além disso, o impacto do borrimento das imagens no desempenho do clustering foi analisado.

Keywords— Processamento de imagens, Células blásticas, Células não-blásticas, clustering, k-means, silhouette_score

I. INTRODUÇÃO

A detecção e classificação clara de células é essencial para diagnosticar e tratar doenças. A análise automatizada de imagens celulares pode ajudar a identificar e diferenciar diferentes tipos de células e, assim, tomar decisões diagnósticas. Neste trabalho, exploramos o impacto de diferentes espaços de cor (RGB e HSV) e o número de clusters (k) na qualidade do agrupamento, medida pelo silhouette score. Além disso, investigamos como o desfoque das imagens pode afetar o desempenho do clustering feito pelo algoritmo k-means.

II. DESCRIÇÃO DA BASE DE DADOS ALL-IDB2

O ALL-IDB2 é uma base de dados utilizada para pesquisa e desenvolvimento em técnicas de processamento e análise de imagens relacionadas à leucemia. Ela fornece um conjunto de imagens microscópicas, que ajudam a identificar e distinguir células blásticas. As imagens são rotuladas com base na presença de células blásticas. As células são classificadas em duas categorias: Células Blásticas ($Y = 1$) indicativas

de leucemia e Células Não-Blásticas ($Y = 0$) são células normais de indivíduos saudáveis. A base de dados ALL-IDB2 é usada para treinar e testar algoritmos de segmentação e classificação de células.

III. PRÉ-PROCESSAMENTO DE IMAGENS

O pré-processamento de imagens é uma etapa crucial na preparação dos dados para análise e modelagem. Para o conjunto de imagens do ALL-IDB2, o pré-processamento foi feito em duas principais operações: alteração do espaço de cor e aplicação de desfoque.

A. Alteração do Espaço de Cor

As imagens que tinham seu espaço de cor original em RGB foram convertidas para HSV visando facilitar a análise de características específicas das células. A função utilizada no processo de alteração do espaço de cor foi `convert_to_hsv`.

B. Aplicação de Desfoque

O objetivo foi suavizar as imagens para reduzir o ruído e melhorar a performance do algoritmo k-means. Os borrimentos utilizados para teste foram 11x11 e 13x13 utilizando a função `apply_blur`.

IV. MÉTRICAS DE AVALIAÇÃO

As métricas de avaliação são importantes para medir a qualidade e a eficácia dos resultados de clustering. Isso ajuda a determinar quão bem os clusters formados refletem a estrutura dos dados e quão distintos são os grupos resultantes. A métrica de avaliação utilizada para medir o desempenho do clustering foi o Silhouette Score. O Silhouette

Score é calculado com base na diferença entre a média das distâncias dentro do cluster e a média das distâncias ao cluster mais próximo.

$$S = \frac{b - a}{\max(a, b)}$$

onde:

- a é a distância média entre um ponto e todos os outros pontos no mesmo cluster.
- b é a distância média entre um ponto e todos os pontos no próximo cluster mais próximo.

Assim, um silhouette score próximo de 1 indica que o ponto está distante de outros clusters e é um bom cluster. Um score próximo de 0 sugere que o ponto está na fronteira entre dois clusters. Quando o score é menor que 0, isso indica que o ponto pode estar agrupado incorretamente.

V. EXPERIMENTOS E COMPARAÇÕES

Nesta seção, descrevemos experimentos usados para avaliar o desempenho do algoritmo k-means na segmentação de imagens de células blásticas e não-blásticas. O algoritmo k-means foi executado com diferentes valores de k (número de clusters), além da conversão das imagens para os espaços de cor RGB e HSV, bem como a aplicação de borramento para simular diferentes condições.

A. Configurações do Algoritmo K-Means

O algoritmo k-means foi executado com dois diferentes valores de k (número de clusters) para determinar a configuração ótima para a segmentação das imagens. Foram utilizados valores de k de 2 e 3, tanto para imagens em RGB quanto em HSV, e k de 2 e 3 para diferentes níveis de borramento (11x11 e 13x13).

B. Métricas de Avaliação

O Silhouette Score avalia a coesão e a separação dos clusters. Pontos com silhouette scores próximo de 1 indicam uma boa separação dos clusters, enquanto scores próximo de 0 ou negativos indicam possíveis problemas na formação dos clusters.

VI. RESULTADOS

Os resultados dos experimentos para segmentação de imagens de células blásticas e não-plásticas foram avaliados com base em diferentes configurações do algoritmo k-means. A seguir estão os principais achados:

A. Resultados para Células Não-Blásticas

- Melhor Modelo de Cor: HSV
- Melhor k : 2
- Média do Silhouette Score para Borramento 11x11: 0.7078
- Média do Silhouette Score para Borramento 13x13: 0.7048

B. Resultados para Células Blásticas

- Melhor Modelo de Cor: RGB
- Melhor k : 2
- Maior Silhouette Score para Borramento 11x11: 0.7435
- Maior Silhouette Score para Borramento 13x13: 0.7378

Conclusão

A análise dos resultados demonstrou diferenças significativas na eficácia do clustering de imagens de células blásticas e não-blásticas, dependendo do modelo de cor e do número de clusters utilizado. A comparação entre diferentes níveis de borramento revelou que, embora o impacto do borramento tenha sido menor para as imagens não-blásticas, ele foi mais pronunciado para as imagens blásticas. Em suma, o uso do modelo de cor e número de clusters apropriados, juntamente com a consideração do impacto do borramento, pode melhorar a precisão da classificação de células blásticas e não-blásticas.

REFERÊNCIA

- [1] <https://medium.com/towards-data-science/semantic-segmentation-of-remote-sensing-imagery-using-k-means-e4c165d9218e>
- [2] https://github.com/alexxyzro/medium/blob/main/Medium_k_Means.ipynb
- [3] https://www.researchgate.net/publication/283185016_Image_Segmentation_Using_K-means_Clustering_Algorithm_and_Subtractive_Clustering_Algorithm
- [4] <https://www.ijert.org/exploring-the-use-of-silhouette-score-in-k-means-clustering-for-image-segmentation>