Bootcamp: Machine Learning

Squad 06: Segmentação de Pessoas | *Tiktok dances dataset*

Atividade 04: Etapa 03

Alunos: Marcus Cabral, Hosana Fernandes Gomes, Francisco Thiago Barbosa, Gisele dos Santos da Silva

Julho / 2025

Sumário

1. Introdução

O Processamento Digital de Imagens (PDI) é uma ferramenta computacional que vem sendo utilizado em ampla escala para extração de informações e manipulação de categorias, conforme as necessidades da área do conhecimento em que é aplicado. Essa metodologia, apesar de mais rápida que um processamento analógico, requer conhecimento especializado e elevada capacidade de processamento computacional, fatores que podem representar limitações em determinadas situações.

Ainda assim, o PDI tem se mostrado muito útil na detecção de padrões e classificação, com aplicações em áreas como: medicina, paleontologia, planejamento urbano e governança. Nesse sentido, este trabalho combina dois notebooks para aperfeiçoar o treinamento de uma Rede Neural Convolucional (CNN) do tipo U-Net para realizar a segmentação de pessoas. A U-Net é uma CNN amplamente usada em segmentação de imagens, que possui um encoder para extração de características com downsampling e um decoder que reconstrói a imagem segmentada via upsampling, permitindo resultados precisos.

Assim, este estudo busca aprimorar a acurácia do notebook existente e investigar aplicações práticas desse processo no contexto da indústria criativa. Ainda, este trabalho pode ser utilizado como ponto de partida para responder à seguinte pergunta de negócio: *"Como a segmentação automática de dançarinos pode ajudar agências de influenciadores a melhorar a curadoria, edição e análise de performance de conteúdos de vídeo de dança?"*. Diante disso, busca-se discutir o potencial do PDI como ferramenta estratégica para otimizar a produção e avaliação de conteúdos audiovisuais voltados para as redes sociais.

2. Metodologia

O dataset utilizado neste trabalho contém cerca de 7600 imagens, dentre as quais 2560 são *prints* de vídeos do tiktok, que contém ampla variedade de corpos em movimento e em diferentes contextos e iluminações. As imagens foram pré-processadas para servir de base para o treinamento da CNN e para a validação do modelo de segmentação de imagens, com foco em destacar o contorno do corpo do fundo, ignorando os ruídos presentes.

Na etapa de pré-processamento, verificou-se a integridade dos arquivos, consistência dos metadados, distribuição de classes e verificação de duplicatas com o auxílio de bibliotecas distintas. A biblioteca **OpenCV** foi utilizada para a leitura de imagens e detecção de objetos segundo parâmetros pré-definidos. Em seguida, a fim de identificar e evitar duplicações ou variações mínimas que pudessem comprometer a generalização do modelo, foi utilizado o algoritmo **Perceptual Hash (pHash)**, para detectar similaridades entre imagens com base nas suas características visuais.

Para a construção da CNN, o modelo u-net deste trabalho foi adaptado de um notebook do kaggle pré-existente, que tinha como principal objetivo estabelecer uma segmentação precisa entre o contorno das pessoas dançando e o fundo das imagens. Adiante, fez-se uso do **TensorFlow**, implementando uma Rede Neural Convolucional do tipo **U-Net** para treinar o modelo na detecção de imagem e fundo. Por fim, essa etapa foi avaliada o com o uso de validação cruzada e por meio de métricas quantitativas específicas de segmentação de imagens, como **IoU (Intersection over Union)**, **Dice Coefficient** e **Accuracy**, além da função de perda **Binary Crossentropy**.

3. Resultados

Por se tratar de um dataset extenso, com alto grau de variabilidade, o algoritmo necessita de várias etapas (*epochs*) para um aprendizado preciso (ver quadro 01). Foram utilizadas 20 épocas, que resultaram em coeficientes favoráveis para uma detecção precisa entre figura e fundo (ver tabela 01). Entretanto, as imagens geradas pela predição do algoritmo apontam para a necessidade de maior treinamento, de modo que o modelo seja capaz de deletar ruídos sem deixar de detectar pontos sensíveis na imagem.

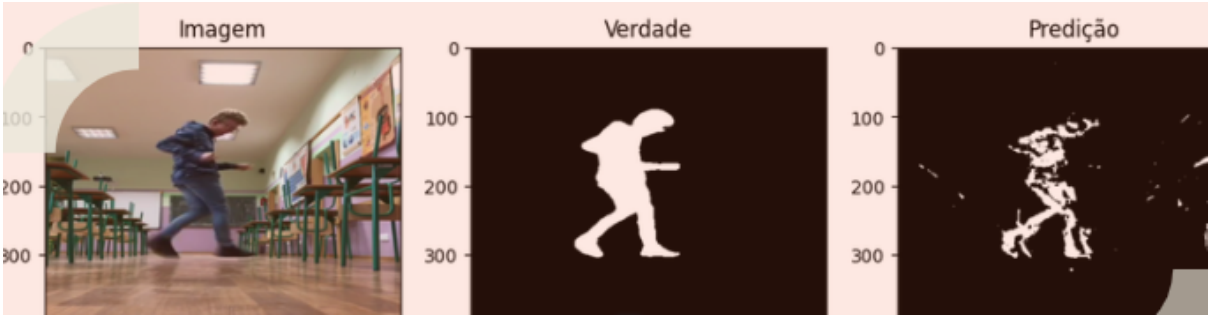
Quadro 01: construção do algoritmo de treinamento[1]

```
model = unet_model()
model.compile(optimizer=Adam(), loss='binary_crossentropy',
metrics=['accuracy'])

train_df,val_df = train_test_split(df, test_size=0.1,random_state=42)
train_gen = DataGenerator(train_df['image_path'].values,
train_df['mask_path'].values, batch_size=8)

val_gen = DataGenerator(val_df['image_path'].values,
val_df['mask_path'].values, batch_size=8)

history = model.fit(train_gen, validation_data=val_gen,epochs=20)
```

Fonte: os autores (2025).

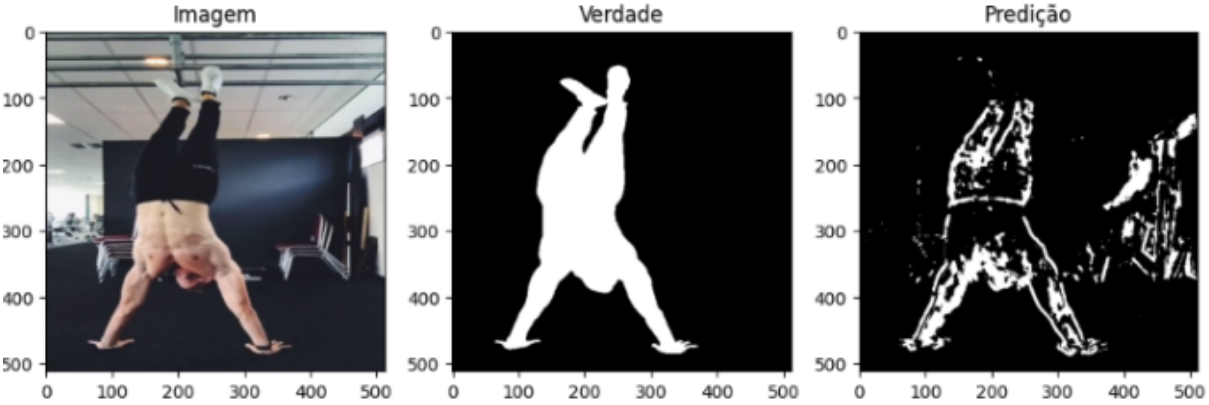
[1] Obs.: o código completo pode ser acessado no google colab.

Figura 01



Fonte: os autores (2025)

Figura 02



Fonte: os autores (2025)

Tabela 01: métricas de avaliação do modelo

Loss	Accuracy	Dice Score	IoU Score
0.0747	0.9716	0.8424	0.7277

Fonte: os autores (2025)

Os dados da tabela 01 revelam que o algoritmo implementado segmenta de maneira satisfatória o corpo do indivíduo em relação ao fundo. Os índices IoU e Dice indicam que, aproximadamente, 73% da área de predição está correta e que 84% das vezes, a máscara prevista se sobrepõe à imagem original, respectivamente. Tais resultados sugerem uma boa generalização, mas que carece de ajustes mais finos para apresentar índices maiores que 0.9. Métricas acima desse valor de referência refletem um modelo cuja segmentação é altamente precisa, em que as máscaras previstas são bastante semelhantes às originais. Isso reduz significativamente a ocorrência de falsos positivos, o que evita a detecção incorreta de elementos do fundo como parte do corpo do indivíduo, como pode ser observado nas figuras 01 e 02.

4. Conclusão

Neste trabalho, foi possível analisar o funcionamento de rede neural convolucional do tipo U-Net para realizar a segmentação automática do corpo de indivíduos em imagens extraídas de vídeos do tiktok, focando na separação entre corpo e fundo. A abordagem envolveu a preparação de um dataset com cerca de 2.600 imagens, a definição de uma arquitetura adequada e o treinamento do modelo utilizando métricas de avaliação como *Loss*, *Accuracy*, *Dice Score* e *IoU*. Os resultados obtidos indicam desempenho satisfatório, com sobreposição média de 84% entre as máscaras previstas e as reais, e 73% de acerto na área segmentada.

Ainda assim, para aprimorar esses resultados, estratégias como a aplicação de *data augmentation* também na etapa de teste, o uso de *early stopping* por meio de *callbacks*, e a experimentação com diferentes arquiteturas disponíveis em notebooks da plataforma Kaggle podem ser exploradas. No entanto, é importante considerar que tais melhorias exigem alta capacidade de processamento computacional, o que pode inviabilizar o treinamento a depender das configurações de hardware do equipamento utilizado.

Por fim, em relação à pergunta de negócio inicial: “Como a segmentação automática de dançarinos pode ajudar agências de influenciadores a melhorar a curadoria, edição e análise de performance de conteúdos de vídeo de dança?”, a segmentação do dataset pode ser utilizada pelas agências de publicidade na identificação das categorias de dança. Para isso, teria de ser feito um treinamento específico para o reconhecimento dos estilos de dança, bem como um ajuste nas classes das imagens. Quanto à edição, o modelo pode ser ajustado para agrupar movimentos semelhantes, o que poderia revelar tendências que impactam positivamente no engajamento dos vídeos.

Em síntese, este trabalho demonstrou o potencial da U-Net como ferramenta eficaz para segmentação automática de corpos em vídeos, apresentando resultados promissores. Apesar do desempenho satisfatório foram identificadas possibilidades de aprimoramento para adaptar o modelo em razão das necessidades das agências de influenciadores. Ao ajustar o dataset para reconhecer estilos de dança e padrões de movimento, é possível utilizar a segmentação de imagens como uma aliada estratégica na curadoria, edição e análise de performance de conteúdos audiovisuais. Desse modo, a integração entre visão computacional e marketing digital revela amplas possibilidades de inovação na produção de vídeos voltados para redes sociais, como o TikTok.

5. Referências

HUYNH, Nghi. *Understanding Evaluation Metrics in Medical Image Segmentation*. Medium blog, 1 mar. 2023. Disponível em: https://medium.com/@nghihuynh_37300/understanding-evaluation-metrics-in-medical-image-segmentation-d289a373a3f. Acesso em: 23 jul. 2025.

KEYLABS. *Best Practices for Image Preprocessing in Image Classification*. Keylabs AI blog, s.d. Disponível em: https://keylabs.ai/blog/best-practices-for-image-preprocessing-in-image-classification/. Acesso em 25 Jul. 2025.

MENINI MATOSAK, Bruno; MEDEIROS, Nilcilene das Graças. *IMGedu.jl – Capítulo 1: Imagem Digital*. Disponível a partir de: https://menimato.github.io/IMGedu.jl/introducao.html#imagem-digital . Acesso em: 26 jul. 2025. Publicado em: 27 ago. 2018.

RSVMUKHESH. *Determining the Number of Epochs*. Medium, 19 mai. 2023. Disponível em: https://medium.com/@rsvmukhesh/determining-the-number-of-epochs-d8b3526d8d06 . Acesso em: 25 jul. 2025.

SHAH, Deval. *Intersection over Union (IoU): Definition, Calculation, Code*. V7 Labs, 30 maio 2023. Disponível em: https://www.v7labs.com/blog/intersection-over-union-guide. Acesso em: 23 Jul. 2025.

SATHEESH, Vishnu. *Hyper Parameter Tuning (GridSearchCV Vs RandomizedSearchCV)*. Analytics Vidhya, 22 dez. 2020. Disponível em: https://medium.com/analytics-vidhya/hyper-parameter-tuning-gridsearchcv-vs-randomizedsearchcv-499862e3ca5. Acesso em: 24 jul. 2025.