

# Transformação de imagens no infravermelho para a faixa do visível aplicada a segmentação semântica de cenas captadas por aeronaves

Bruno N. C. Dantas<sup>1</sup>, Élcio H. Shiguemori<sup>2</sup>  
<sup>1</sup>Instituto Tecnológico de Aeronáutica, <sup>2</sup>Instituto de Estudos Avançados

**RESUMO** – A segmentação semântica de imagens aéreas fornece o significado, ao nível de pixel, do ambiente de voo para auxílio na tomada de decisão em sistemas de navegação autônoma. As redes neurais convolucionais de aprendizado supervisionado demandam grande esforço de rotulagem de imagens antes do voo, pois as mudanças de parâmetros da câmera ou das características do local podem impactar nos resultados de inferência. Para minimizar esse esforço, esse trabalho emprega redes adversárias generativas para tradução de imagens no domínio do infravermelho para o domínio do espectro do visível com vistas a gerar imagens sintéticas a serem segmentadas pela rede neural segmentadora treinada no espectro visível.

## I - INTRODUÇÃO

**Motivação:** Para a aplicação de Redes neurais artificiais que segmentam imagens com uso de aprendizado supervisionado é necessário a criação de um banco de dados com as máscaras de segmentação que indique as classes de superfícies como estrada, construções e outros de uma porção da área a ser segmentada[1]. O processo de criação de máscaras de segmentação é custoso e dificulta a implantação desse tipo de sistema.

**Objetivo:** Este trabalho busca analisar a eficiência da técnica de transformação de imagens para adaptação de domínio, realizada pelas redes adversárias generativas (GANs), para auxílio na segmentação semântica de imagens aéreas no espectro do infravermelho visando à navegação autônoma por imagens, que utiliza um modelo treinado em imagens no espectro do visível de uma mesma região de voo, aliviando assim o esforço de criação novas máscaras para o infravermelho.

**Estratégias abordadas no trabalho:** Foram realizados voos com drone phantom 4 equipado com uma câmera no espectro do visível (RGB) de alta resolução e uma câmera multiespectral Sequoia, como ilustrado na figura 1. Deste voo, foram selecionadas 253 imagens capturadas pela câmera no espectro do visível e 253 imagens no infravermelho obtidas pela câmera multiespectral. A figura 2 destaca duas cenas captadas pelas diferentes câmeras.



Figura 1. Drone Phantom 4 e câmera multiespectral Sequoia



Figura 2. Imagens capturadas pelas câmeras durante o voo.

## II- MÉTODOS

### Rede neural para segmentação semântica

**Banco de dados utilizado na rede segmentadora:** Todas as imagens foram captadas na mesma altitude (80m). As máscaras foram manualmente rotuladas pelo software Labelme. Os pares de imagens e máscaras foram divididos com os seguintes quantitativos: 213 para treinamento, 21 para validação e 19 para teste.

**Modelo utilizado para segmentação:** Rede Deeplabv3[2] pré-treinada com o banco de dados COCO[3].

**Classes empregadas:** asfalto, construção e carro.



Figura 3. Rede para segmentação semântica

### Rede neural adversária generativa

**Banco de dados utilizado para treinamento:** Todas as imagens obtidas tanto pela câmera no espectro do visível quanto pela banda do infravermelho capturada pela câmera multiespectral.

**Modelo utilizado para a transformação das imagens do infravermelho para o espectro do visível:** A rede adversária generativa selecionada foi a CycleGAN, que realiza a tradução de imagens de maneira não-supervisionada através de consistência de ciclo sem que as imagens tenham que estar pareadas de alguma forma[4].

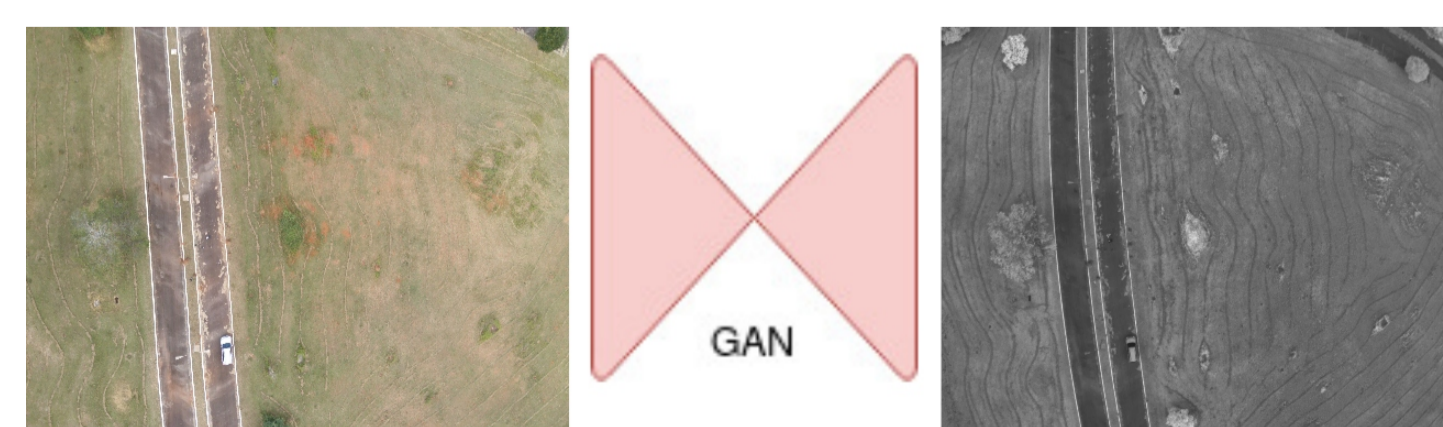


Figura 4. Rede adversária generativa

### Metodologia utilizada

Após a fase de treinamento de cada rede neural feita de maneira separada:

- 1) Realizou-se a inferência da rede segmentadora nas imagens do visível e do infravermelho.
- 2) Foram geradas imagens sintéticas no espectro do visível a partir das imagens no infravermelho utilizando a rede neural adversária generativa.
- 3) Finalmente, foi feita a inferência da rede segmentadora nas imagens sintéticas.

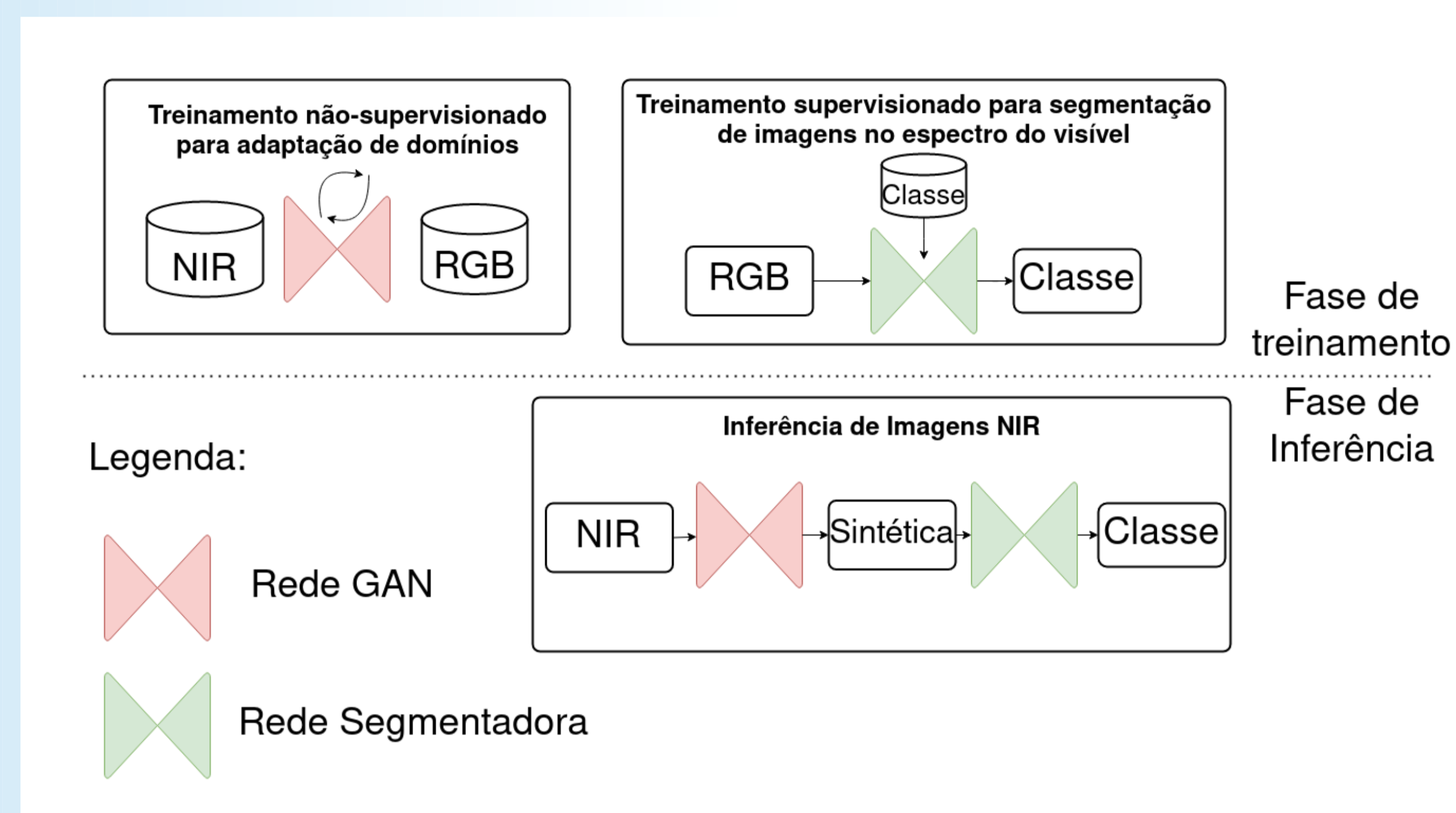


Figura 5. Metodologia

## III- RESULTADOS

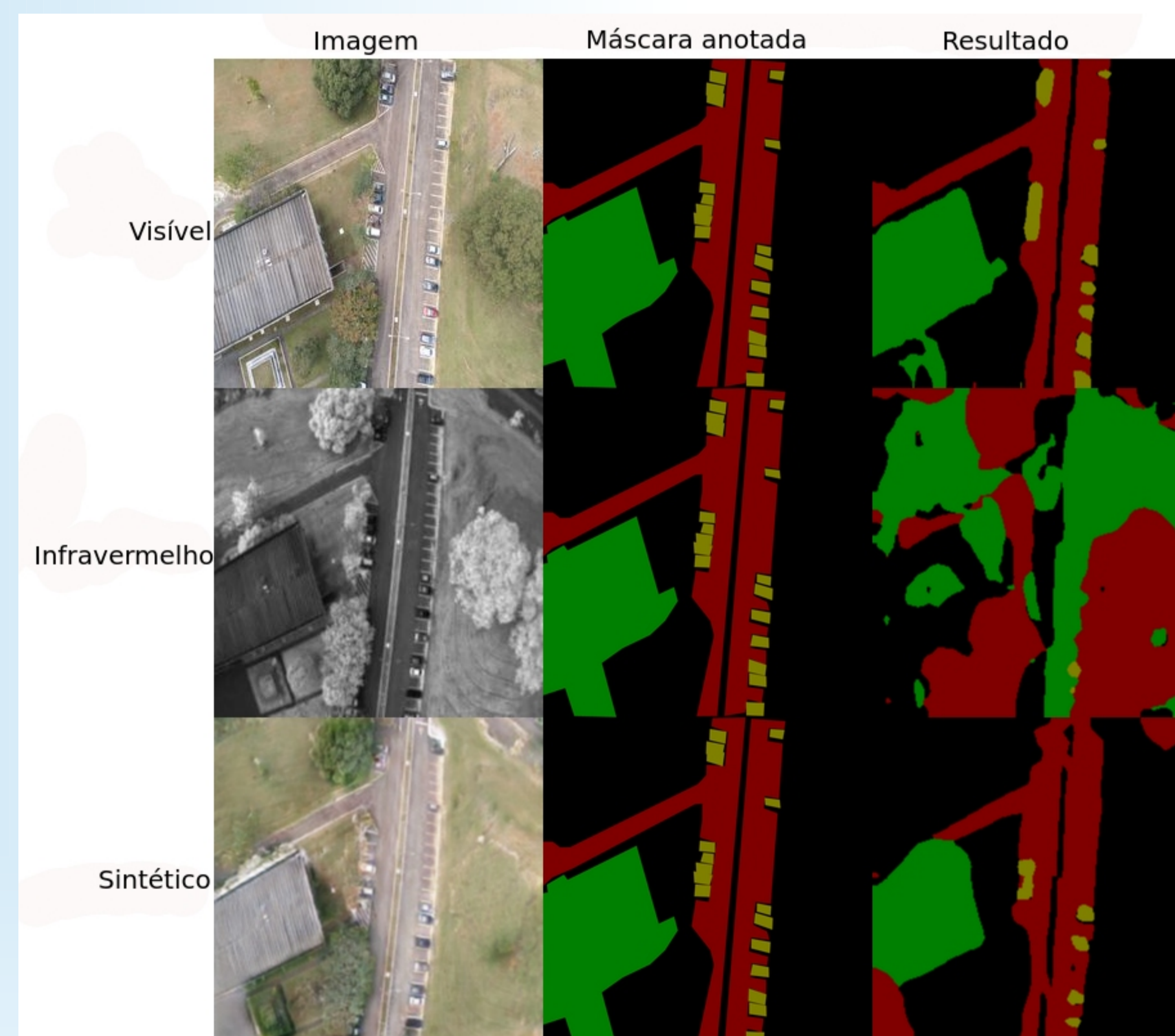


Figura 6. Resultados de segmentação.

Imagem	Acurácia	IOU
Visível	0,97	0,84
Infravermelho	0,18	0,06
Sintética	0,88	0,57

Tabela 1. Resultados de segmentação.

## IV - Conclusão

Foi demonstrado a eficiência na tradução de imagens pela rede adversária generativa para auxiliar na tarefa de segmentação de imagens, diminuindo o esforço de rotulagem de imagens obtidas em outro espectro, como o infravermelho e levando os modelos previamente treinados a bons resultados de inferência com sensores distintos.

## REFERÊNCIAS

- [1] Nassar, A., Amer, K., ElHakim, R., & Elhelw, M. (2018). A Deep CNN-Based Framework For Enhanced Aerial Imagery Registration with Applications to UAV Geolocalization. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1594-159410.
- [2] Chen, L., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. ArXiv, abs/1706.05587
- [3] Lin, T., Maire, M., Belongie, S.J., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C.L. (2014). Microsoft COCO: Common Objects in Context. ECCV
- [4] J. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.