

FUSÃO MULTIESPECTRAL EM DEEP LEARNING: MELHOR DELINEAMENTO E IDENTIFICAÇÃO DE REDE VIÁRIA EM AMBIENTE URBANO

ANTONIO G. GUIMARÃES FILHO¹
MILTON H. SHIMABUKURO²
ALUIR P. DAL POZ³

¹Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia – guimaraes.filho@unesp.br

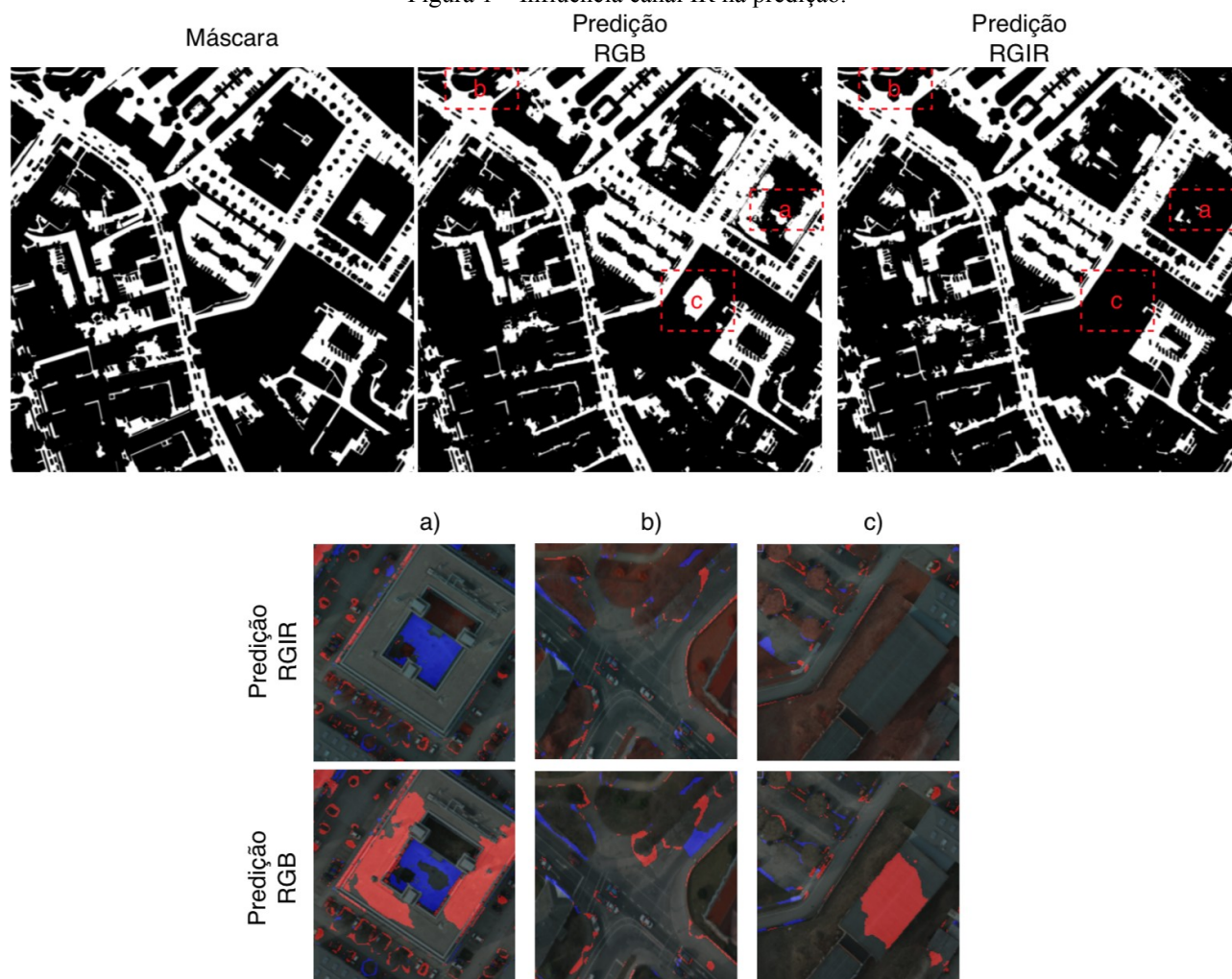
²Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia – milton.h.shimabukuro@unesp.br

³Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia – aluir.dal-poz@unesp.br

O Brasil é um país que possui 1,7 milhões de quilômetros de rodovias distribuídas de forma irregular pelo território nacional[1]. Os diferentes ambientes/contextos influenciam as características das vias. Por exemplo, em termos de dinamismo, a área urbana possui taxas maiores do que a rural. Mudanças de geometria, novas conexões, eliminação de trechos, mudanças de direções são alguns exemplos de como as informações sobre rodovias são relevantes para planejamento e gestão urbana, bem como para a economia. Tal dinamismo e complexidade tornam a obtenção de informações da rede viária, de forma ágil, um grande desafio. Nesse contexto, a extração de vias a partir de imagens de Sensoriamento Remoto é complexa devido a algumas características tais como: tipo de sensor, resolução espectral e espacial, clima, variação de luz e característica do solo. [2] afirmam que, a partir de 2014, o interesse sobre aprendizado profundo (*Deep Learning* - DL) aumentou na comunidade de Sensoriamento Remoto pela capacidade de extrair características a partir dos dados sem intervenção humana para construir regras. Desde então, algoritmos de DL têm obtido notáveis resultados na extração de rede viária a partir de imagens multiespectrais (espectro do visível) [3]. Paralelamente, avanços tecnológicos nas décadas de 80/90 impulsionaram pesquisas sobre técnicas de fusão de dados [4-6]. [7] afirmam que dados de observação da Terra não estão limitados aos sensores ópticos que atuam na faixa do visível. Existem outros sensores e fontes que podem ser utilizadas para complementar a informação de extração de objetos. Portanto, este estudo pretende analisar a influência de dados do Infravermelho (IR) e da escolha da função perda (*Loss*) na extração de rede viária na área urbana por meio de DL. Para isso, o conjunto de dados referência ISPRS Potsdam [8] foi utilizado junto com a arquitetura U-Net, a qual foi escolhida por sua simplicidade de implementação e pela possibilidade de realizar adaptações. Neste trabalho, foram feitas as seguintes alterações na arquitetura: uso de classificadores VGG-16 e ResNet-50, inserção de *Batch Normalization* (BN) para regularização dos valores de mapas de característica, imagens de entrada com tamanho e canais diferentes da original, recortes aleatórios (*random crop*), outras transformações para aumento artificial de dados (espelhamento horizontal, vertical e rotação de 90°), utilização de *Binary CrossEntropy*, *Dice Loss*, *Tversky* ou *Focal Tversky Loss* para o desbalanceamento com a classe não rodovia e utilização de transferência de aprendizado (*Transfer Learning*) de redes treinadas no conjunto ImageNet. O uso de BN visa a normalização da distribuição dos valores obtidos nos filtros a cada subamostra (*batch*) antes da função de ativação. Isso permite que o treinamento alcance a convergência de forma mais ágil e se obtenha maior acurácia na otimização. Enquanto o aumento artificial de dados permite aumentar a generalização do modelo e evita o superajuste aos dados de treinamento. Todas as adaptações foram realizadas de forma a aproveitar as melhores práticas de treinamento das redes convolucionais profundas. O conjunto de dados da cidade de Potsdam, Alemanha, é constituído por 38 recortes de ortofotos, divididos em 24 amostras para conjunto treinamento e 14 para teste, cada recorte possui *Ground Sample Distance* (GSD) de 5 cm e combinação RGB, RG-IR ou RGB-IR com resolução radiométrica de 8 bits por banda e máscaras de segmentação para treinamento e teste de 6 classes, entre as quais encontra-se a de Superfícies Impermeáveis (*Impervious surfaces*). A “via” está incluída nessa classe de superfícies juntamente com outras feições (p.ex: estacionamentos), entretanto apresenta “buracos”, tais como copa de vegetação e carros. Um procedimento de matemática de bandas foi realizado para separação da classe de superfícies e convertê-la em máscara binária. As ortoimagens foram reamostradas para 10cm de modo a inserir contexto da via em recortes de tamanho 600x600 pixels com sobreposição vertical e horizontal de 50% e normalização para valores de pixel entre [0,1]. As redes ResUnet e Unet foram treinadas com diferentes funções de perda, assim como em dados RGB/RGIR por 180 épocas (76.545 iterações), tempo médio de 4 horas de treinamento por modelo em computador equipado com placa GPU NVIDIA RTX 3080 10GB DDR6x, 128 GB RAM, 1TB HD SSD. De acordo com a Tabela 1, modelos otimizados com funções de perda que penalizam falsos negativos (*tversky* e *focal tversky*) proporcionam os valores mais altos na métrica *recall*, por exemplo a Unet_ftv apresenta percentual de 92,10 no conjunto de imagens RGB. Entretanto, houve perda significativa de *precision* (82,62%). As redes otimizadas pelas funções de perda *dice* e *bce+dice* apresentaram os melhores resultados nas métricas, principalmente F1 e IoU. O modelo ResUnet_bce+dice obteve o melhor resultado no conjunto teste de Potsdam com IoU = 81,12 e F1 = 89,52. Em uma mesma arquitetura, a escolha da função de perda para

otimização pode acarretar melhores resultados nas previsões finais. Por exemplo, a ResUnet_bce+dice possui acréscimo de 4% em IoU e 2,3% em F1, assim como o modelo Unet_dice tem 2% em IoU e 1,1% em F1 em relação aos otimizados pela função *tversky*. Ambos os casos no conjunto com dados IR. Outro aspecto está relacionado ao uso da imagem IR no conjunto de dados. Praticamente todos os modelos apresentaram desempenho superior utilizando os dados com IR nas métricas F1 e IoU. Somente os modelos ResUnet_dice e ResUnet_tv apresentaram resultados superiores com os dados RGB. No caso do modelo ResUnet_bce+dice, o acréscimo foi de 1,8% em IoU e 1% em F1 em relação ao treinado com RGB. Com isso, é possível observar que o canal IR agrega informação na separação de classes (especificamente na métrica *precision*), principalmente com a vegetação e na diferença de vias e telhados. Essa separação acarreta um melhor delineamento da via. Sob o aspecto visual, a Figura 1 exibe essa tendência nas previsões. A predição da classe superfícies impermeáveis com IR possui contornos mais definidos (exemplos 1a), 1b) e 1c)) e também diminui a confusão com texturas diferentes (exemplos 1a) e 1c)) frente a extração no espectro do visível. Por fim, os testes realizados mostram que a escolha da função de perda e o uso do IR influenciam na predição das vias e das outras feições dentro da classe de superfícies na área urbana. Embora os modelos treinados com funções de perda diferentes tenham influência no resultado, não foi encontrada uma relação entre a escolha e a predição final. Outros trabalhos na linha de fusão de dados de sensores/fontes e escolha de função *Loss* são relevantes na extração de objetos bem como melhor compreensão da superfície imageada. Implementação e resultados disponíveis em https://github.com/tunofilho/CBCG_2022_article.

Figura 1 – Influência canal IR na predição.



Fonte: Autor (2022). Predições realizadas com entrada RGB e RGIR do modelo de maior desempenho na métrica IoU e F1. Recortes (a), (b) e (c) em azul: falsos negativos; em vermelho: falsos positivos sobrepostos na imagem.

Tabela 1 – Resultado no conjunto teste de ISPRS Potsdam.

Modelos	IoU		Precision		Recall		F1	
	RGB	RGIR	RGB	RGIR	RGB	RGIR	RGB	RGIR
ResUnet_ftv	79,65	79,95	86,79	87,19	90,60	90,54	88,61	88,78
ResUnet_dice	79,85	79,83	88,68	88,94	88,93	88,60	88,73	88,71
ResUnet_tv	79,04	77,95	86,10	83,98	90,58	91,55	88,23	87,53
ResUnet_bce+dice	79,78	81,12	89,40	90,77	88,08	88,39	88,67	89,52
Unet_ftv	77,10	78,47	82,62	84,35	92,10	91,81	87,00	87,88
Unet_dice	79,75	80,03	89,84	89,34	87,67	88,45	88,65	88,83
Unet_tv	78,96	79,24	86,23	85,26	90,34	91,75	88,18	88,33
Unet_bce+dice	78,85	80,01	89,88	88,97	86,61	88,75	88,11	88,81

Fonte: Autor (2022). **Em negrito**: melhores resultados do modelo na métrica por tipo de imagem de entrada. **Em negrito e sublinhado**: melhor resultado na métrica. ftv: focal tversky ; tv: tversky ; bce+dice: binary cross entropy e dice loss

Palavras-chaves: Fusão multispectral; Extração de rodovias; Segmentação semântica; aprendizado profundo.

Referências

- [1] CNT. Pesquisa CNT de rodovias 2019. Brasília: Sest Senat, 2019. 236 p. Disponível em: https://pesquisarodovias.cnt.org.br/Downloads/Edicoes//2019/Relatório%20Gerencial/Relatorio_Gerencial_Pesquisa_CNT_Rodovias_2019.pdf. Acesso em: 08 set. 2020.
- [2] MA, L. et al. ISPRS Journal of Photogrammetry and Remote Sensing Deep learning in remote sensing applications : A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing, v. 152, n. March, p. 166–177, 2019. ISSN 0924-2716. Disponível em: <<https://doi.org/10.1016/j.isprsjprs.2019.04.015>>.
- [3] ABDOLLAHI, A. et al. Deep Learning Approaches Applied to Remote Sensing Datasets for Road Extraction: A State-Of-The-Art Review. Remote Sensing, 2 maio. 2020. v. 12, n. 9, p. 1444. Disponível em: <<https://www.mdpi.com/2072-4292/12/9/1444>>.
- [4] SCHENK, T.; CSATHÓ, B. Fusion of Lidar Data and Aerial Imagery for a More Complete Surface Description. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, v. 34, n. 3/A, p. 310–317, 2002. Disponível em: <http://rsl.geology.buffalo.edu/documents/schenk{_}csatho{_}gr>.
- [5] MENA, J. B. State of the art on automatic road extraction for GIS update: A novel classification. Pattern Recognition Letters, v. 24, n. 16, p. 3037–3058, 2003. ISSN 01678655.
- [6] LIU, Y. et al. Deep learning for pixel-level image fusion: Recent advances and future prospects. Information Fusion, Elsevier, v. 42, n. September 2017, p. 158–173, 2018. ISSN 15662535. Disponível em: <<https://doi.org/10.1016/j.inffus.2017.10.007>>.
- [7] AUDEBERT, N.; SAUX, B. L.; LEFÈVRE, S. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. ISPRS Journal of Photogrammetry and Remote Sensing, v. 140, p. 20–32, 2018. ISSN 09242716.
- [8] ROTTENSTEINER, F. et al. ISPRS Test Project on Urban Classification and 3D Building Reconstruction. ISPRS - Commission III - Photogrammetric Computer Vision and Image Analysis Working Group III / 4 - 3D Scene Analysis, p. 1–16, 2013. ISSN 18770428.