



COLEÇÃO
ESTUDOS
CARIOCAS

Volume
13

Edição
4

*Autor(a) correspondente
jediello.souza@ibge.gov.br

Submetido em 15 nov 2025

Aceito em 13 jan 2026

Publicado em 05 fev 2026

Como Citar?

SOUZA, J. S.; SILVA, A. D.;
NUNES, I. M. Detecção de
favelas cariocas em ortoimagens
utilizando deep learning com
arquitetura U-Net. *Coleção
Estudos Cariocas*, v. 13, n. 4,
2026.
DOI 10.71256/19847203.13.4.195.2025.

O artigo foi
originalmente
submetido em
PORTUGUÊS.

As traduções para
outros idiomas foram
revisadas e validadas
pelos autores e pela
equipe editorial. No
entanto, para a
representação mais
precisa do tema
abordado,
recomenda-se que os
leitores consultem o
artigo em seu idioma
original.



Detecção de favelas cariocas em ortoimagens utilizando deep learning com arquitetura U-Net

Detection of Rio de Janeiro favelas in orthoimages using deep learning with the U-Net architecture

Detección de favelas de Río de Janeiro en ortoimágenes utilizando deep learning con la arquitectura U-Net

Jediello Sales de Souza¹, Andrea Diniz da Silva² e Ian Monteiro Nunes³

¹ Instituto Brasileiro de Geografia e Estatística, Superintendência Estadual do Rio de Janeiro, Av. Beira Mar, 436 - Centro, Rio de Janeiro/RJ, CEP 20021-060, ORCID: 0009-0009-6053-7065, jediello.souza@ibge.gov.br

² Instituto Brasileiro de Geografia e Estatística, Escola Nacional de Ciências Estatísticas, Rua André Cavalcanti, 106 - Santa Teresa, Rio de Janeiro/RJ, CEP 20231-050, ORCID: 0000-0001-9116-0162, andrea.diniz@ibge.gov.br

³ Instituto Brasileiro de Geografia e Estatística, Diretoria de Pesquisas, Rua Pacheco Leão, 1235 - Jardim Botânico, Rio de Janeiro/RJ, CEP 22460-905, ORCID: 0000-0003-3445-4169, ian.nunes@ibge.gov.br

Resumo

Este estudo avaliou o desempenho da arquitetura U-Net na identificação de favelas em ortoimagens de alta resolução, utilizando máscaras manuais como referência. Foram comparados modelos treinados com e sem data augmentation. O modelo com data augmentation apresentou melhor desempenho em IoU, F1-Score e Precisão, enquanto o modelo sem *augmentation* obteve maior Revocação, evidenciando o trade-off entre sensibilidade e controle de falsos positivos. Apesar das dificuldades em áreas pequenas e de baixo contraste visual, os resultados confirmam o potencial da U-Net para o mapeamento de assentamentos precários.

Palavras-chave: favela, ortoimagem, deep learning, planejamento urbano, geotecnologias

Abstract

This study evaluated the performance of the U-Net architecture for identifying informal settlements in high-resolution orthoimagery, using manually annotated masks as reference. Models trained with and without data augmentation were compared. The model with data augmentation achieved better performance in terms of IoU, F1-Score, and Precision, while the model without augmentation obtained higher Recall, highlighting the trade-off between sensitivity and false-positive control. Despite challenges related to small areas and low visual contrast, the results confirm the potential of U-Net for mapping informal settlements.

Keywords: favela, orthoimage, deep learning, urban planning, geotechnologies

Resumen

Este estudio evaluó el desempeño de la arquitectura U-Net en la identificación de asentamientos informales en ortoimágenes de alta resolución, utilizando máscaras anotadas manualmente como referencia. Se compararon modelos entrenados con y sin data augmentation. El modelo con aumento de datos presentó un mejor desempeño en términos de IoU, F1-Score y Precisión, mientras que el modelo sin *augmentation* obtuvo una mayor Revocación, evidenciando el trade-off entre sensibilidad y control de falsos positivos. A pesar de las dificultades asociadas a áreas pequeñas y de bajo contraste visual, los resultados confirman el potencial de la U-Net para el mapeo de asentamientos informales.

Palabras clave: favela, ortomagen, deep learning, planificación urbana, geotecnologías

1 Introdução

O crescimento urbano acelerado e muitas vezes desordenado impõe desafios crescentes à gestão das cidades, comprometendo o desenvolvimento sustentável. Segundo o Relatório Mundial das Cidades (ONU BRASIL, 2022), a população urbana mundial deve passar de 56% em 2021 para 68% em 2050.

No Brasil, a urbanização esteve historicamente associada à expansão de assentamentos precários, resultado da ausência de políticas públicas eficazes e da falta de acesso à moradia e infraestrutura adequadas. No Rio de Janeiro, o processo de remoção de cortiços no final do século XIX levou famílias a ocupar morros e áreas periféricas, originando as primeiras favelas (Ling, 2018; Marins, 1998).

Diante da escala global da urbanização, a Agenda 2030 da ONU estabeleceu os 17 Objetivos de Desenvolvimento Sustentável (ODS), entre os quais o ODS 11 – Cidades e Comunidades Sustentáveis visa “tornar as cidades inclusivas, seguras, resilientes e sustentáveis”, prevendo a melhoria e urbanização das favelas até 2030 (ONU BRASIL, 2023). Um dos principais desafios para atingir essa meta é a identificação e o monitoramento contínuo de assentamentos precários, essenciais para o planejamento urbano e a formulação de políticas públicas.

Tradicionalmente, essas informações são obtidas pelos censos demográficos, realizados a cada dez anos, o que limita o acompanhamento de transformações no período intercensitário. Nesse contexto, o uso de imagens de satélite e técnicas de inteligência artificial surge como uma alternativa promissora, permitindo identificar, mapear e monitorar favelas de forma mais ágil e precisa.

O sensoriamento remoto tem sido amplamente empregado para a identificação e o monitoramento de assentamentos informais, aproveitando imagens de satélite de diferentes resoluções espaciais e temporais em contextos urbanos diversos. Estudos demonstram que características espectrais, texturais, geométricas e morfológicas extraídas dessas imagens permitem distinguir áreas formais e informais, especialmente quando combinadas com técnicas de processamento digital de imagens e aprendizado de máquina (Kemper *et al.*, 2015; Alrasheedi *et al.*, 2021; Cinnamon; Noth, 2023).

Abordagens tradicionais, como a análise baseada em objetos e algoritmos de aprendizado de máquina clássico, apresentam bons resultados, porém demandam extração e seleção manual de características e ajustes específicos para cada contexto urbano (Ghaffarian; Emtehani, 2021; Oliveira *et al.*, 2023). Nos últimos anos, métodos baseados em aprendizado profundo, especialmente redes neurais convolucionais e modelos de segmentação semântica, têm se consolidado como o estado da arte na análise de imagens de satélite, demonstrando maior capacidade de capturar a complexidade morfológica e a heterogeneidade espacial características das favelas (Wurm *et al.*, 2019; Maiya; Babu, 2018; Lu *et al.*, 2021; Abascal *et al.*, 2022). Essas abordagens ampliam o potencial de mapeamento e monitoramento contínuo de assentamentos informais, fornecendo subsídios mais precisos para o planejamento urbano e a formulação de políticas públicas.

O presente trabalho busca contribuir para esse esforço, aplicando modelos de aprendizado profundo em imagens de satélite para a identificação e o monitoramento da dinâmica de favelas na cidade do Rio de Janeiro.

2 Materiais e Métodos

2.1 Identificação e Delimitação das Áreas de Interesse

Foram utilizadas no estudo ortoimagens RGBI com resolução espacial de 15 cm (Ground Sample Distance - GSD), adquiridas durante voo fotogramétrico realizado em abril de 2024 pela câmera aerofotogramétrica UltraCam Osprey 4.1. Os dados foram cedidos pelo Instituto Municipal de Urbanismo Pereira Passos (IPP), órgão da

Prefeitura da Cidade do Rio de Janeiro. As ortoimagens são compostas pelas bandas espectrais do vermelho (Red), verde (Green), azul (Blue) e infravermelho próximo (Near Infrared - NIR).

As imagens encontram-se no sistema de referência de coordenadas EPSG:31983, que corresponde ao sistema geodésico SIRGAS 2000 associado à projeção cartográfica Universal Transversa de Mercator (UTM), fuso 23 Sul (meridiano central 45°W). Este sistema de referência é adequado à área de estudo, uma vez que o município do Rio de Janeiro está localizado próximo ao meridiano de secância do fuso, o que contribui para a minimização das distorções lineares e areais nas análises espaciais.

Para este estudo, foram utilizadas exclusivamente as bandas do espectro visível (RGB). Considerando o elevado custo computacional associado ao processamento de imagens de grande extensão territorial, foi elaborada uma grade espacial composta por quadrados de 512 × 512 metros, delimitada pela área do município do Rio de Janeiro.

Com o objetivo de reduzir a resolução espacial das imagens, originalmente de 15 cm, foi realizada a reamostragem dos pixels pelo método bilinear, resultando em uma resolução de 50 cm. Essa nova resolução preserva uma adequada capacidade de identificação dos componentes urbanos, ao mesmo tempo em que reduz significativamente o custo computacional do processamento.

Os quadrados localizados em ilhas desabitadas foram descartados, enquanto aqueles situados na fronteira municipal foram mantidos apenas quando pelo menos um terço de sua área se encontrava dentro dos limites do município.

Ao final desse processo, a grade resultante passou de 5.002 para 4.647 quadrados. Esses quadrados foram, então, utilizados para recortar a imagem do município do Rio de Janeiro (com resolução espacial de 1 m), gerando múltiplas imagens menores e, portanto, mais viáveis para análise computacional.

A cidade do Rio de Janeiro, conforme definido em seu Plano Diretor de Desenvolvimento Urbano Sustentável (PDUS), é subdividida em diferentes unidades territoriais destinadas ao planejamento e controle do desenvolvimento urbano. Entre essas unidades estão as Áreas de Planejamento (APs), cinco áreas definidas a partir de critérios de compartimentação ambiental, características históricas e padrões de uso e ocupação do solo (Rio de Janeiro, 2025):

- Área de Planejamento 1 (AP1) – corresponde à região central da cidade;
- Área de Planejamento 2 (AP2) – abrange a Zona Sul e Grande Tijuca;
- Área de Planejamento 3 (AP3) – compreende a Zona Norte, com exceção da Grande Tijuca e da Região Administrativa VII (São Cristóvão);
- Área de Planejamento 4 (AP4) – corresponde à Região Sudoeste, criada mais recentemente;
- Área de Planejamento 5 (AP5) – engloba os bairros da Zona Oeste.

Com o objetivo de assegurar uma representatividade territorial abrangente, optou-se por utilizar as Áreas de Planejamento definidas oficialmente pela Prefeitura. Essa decisão considera que as diferenças físicas e socioeconômicas da população exercem influência direta sobre a configuração e as características do espaço urbano.

Assim, cada quadrado da grade foi associado a uma Área de Planejamento (AP) do município (Figura 1), conforme a divisão oficial disponibilizada pela plataforma DATA.RIO (2023). Nos casos em que um quadrado se sobrepunha a duas ou mais APs, foram estabelecidos critérios específicos de atribuição, assegurando a consistência da classificação territorial:

1. Quadrados contendo regiões classificadas como favelas pelo IBGE ou pela Prefeitura do Rio foram atribuídos à AP que englobasse essa área classificada;
2. Quadrados sem áreas de favela, ou contendo áreas de favela em mais de uma AP, foram designados à AP que apresentasse a maior extensão territorial dentro do quadrado.

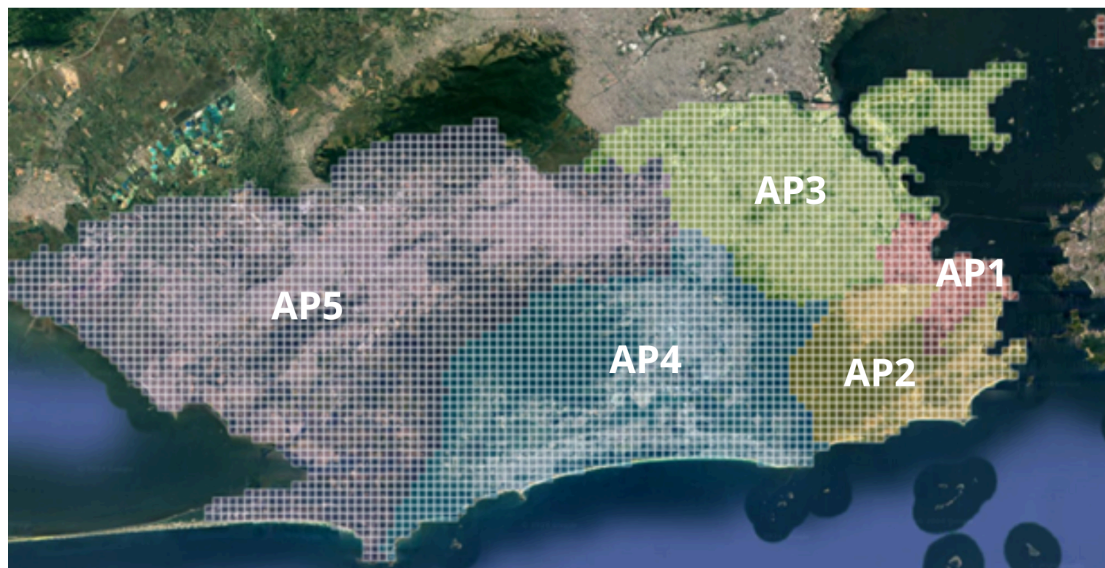


Figura 1: Grade de quadrados de 512 m² aplicada sobre a imagem da cidade do Rio de Janeiro. A malha foi utilizada para segmentar a área urbana em recortes menores e padronizados, facilitando o processamento das imagens de satélite. Distribuída pelas Áreas de Planejamento da Prefeitura do Rio de Janeiro.

Fonte: Elaboração própria

A identificação dos quadrados da grade que continham as áreas de interesse foi realizada com base em dados espaciais vetoriais em formato shapefile. Esses dados correspondem à malha de setores censitários do Censo 2022, disponibilizada pelo IBGE, inicialmente classificados como Aglomerados Subnormais (AGSN), denominação posteriormente atualizada para Favelas e Comunidades Urbanas (FCU).

A mudança de nomenclatura reflete um processo de revisão conceitual conduzido pelo IBGE, em diálogo com movimentos sociais, a comunidade acadêmica e outros órgãos governamentais. Essa atualização busca adotar uma terminologia mais adequada e respeitosa, substituindo a expressão “Aglomerados Subnormais”, que vinha sendo considerada inadequada por reforçar estigmas sociais. No entanto, é importante destacar que a alteração foi apenas terminológica, não havendo mudanças nos critérios técnicos utilizados para a identificação e o mapeamento dessas áreas (IBGE, 2024).

Complementarmente, foram incorporados dados do Instituto Municipal de Urbanismo Pereira Passos (IPP). Esses dados, disponibilizados no portal DATA.RIO (2019), delimitam as áreas oficialmente reconhecidas como favelas.

Para o IBGE, a caracterização de um aglomerado subnormal ocorre quando há ocupação irregular da terra, associada a pelo menos uma das seguintes condições (IBGE, 2020):

- Precariedade de serviços públicos essenciais, como abastecimento de água, fornecimento de energia elétrica, coleta de lixo ou esgotamento sanitário;
- Padrão urbanístico irregular, refletido na presença de vias de circulação estreitas, alinhamento irregular, lotes de tamanhos e formas desiguais, ausência de calçadas ou construções não regularizadas por órgãos públicos;
- Restrição à ocupação do solo, quando os domicílios se encontram em áreas

protegidas por legislação ambiental, faixas de domínio de rodovias ou ferrovias, áreas contaminadas, entre outras situações de uso inadequado do solo urbano.

Além disso, após a identificação e delimitação das áreas, o IBGE associa os aglomerados subnormais às unidades operacionais do Censo Demográfico, denominadas setores censitários. Cada setor deve compreender áreas contíguas com, no mínimo, 51 domicílios, respeitando limites político-administrativos e garantindo a coerência territorial e operacional da coleta censitária (IBGE, 2020).

Por sua vez, o IPP — órgão vinculado à Prefeitura do Rio de Janeiro — utiliza o termo favela com um enfoque socioespacial e administrativo, buscando reconhecer e mapear territórios consolidados no tecido urbano. A definição adotada pelo IPP leva em conta um conjunto mais amplo de características, que incluem tanto aspectos físicos quanto sociais e cadastrais. Segundo o Instituto, uma área é considerada favela quando apresenta (Rio de Janeiro, 2012):

1. Ocupação irregular da terra;
2. Ausência de títulos de propriedade formais (o que não implica ilegalidade da ocupação);
3. Tecido urbano disposto de forma irregular;
4. Lotes pequenos e indefinidos;
5. Vias estreitas;
6. Infraestrutura de saneamento precária;
7. Equipamentos sociais inexistentes ou insuficientes;
8. Habitações precárias e em desacordo com as normas urbanísticas;
9. Inexistência de normas urbanísticas especiais aplicáveis à área;
10. Não inserção dos imóveis nos cadastros imobiliários municipais;
11. Predominância de população de baixa renda.

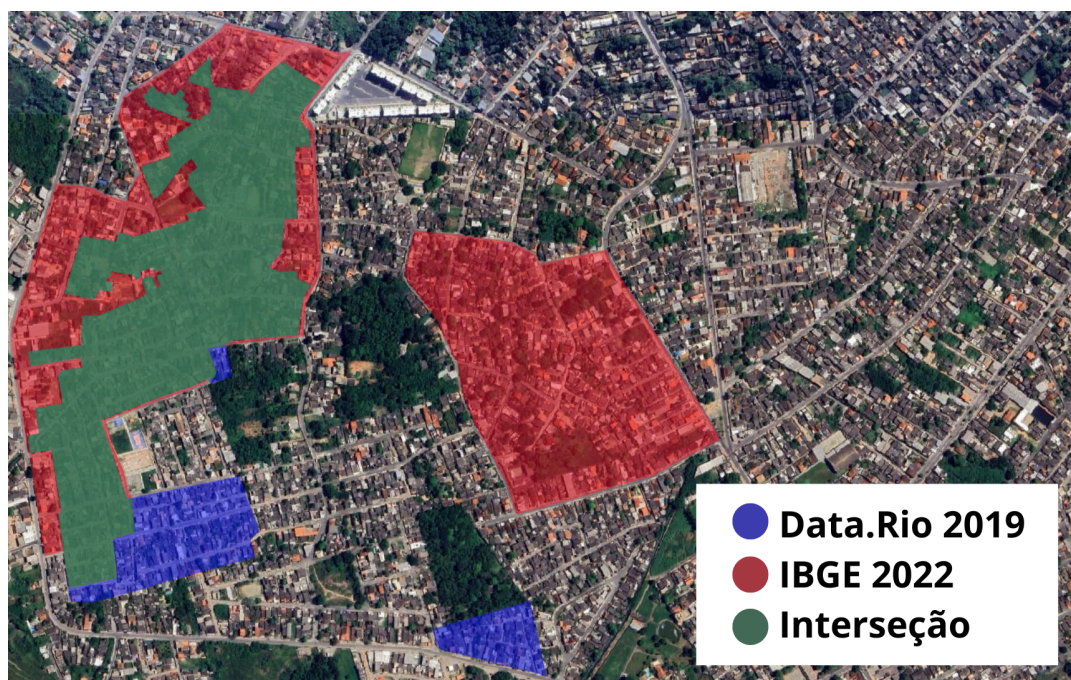


Figura 2: Áreas de favelas classificadas pelo IBGE e pelo Data.Rio. A sobreposição das duas bases permite visualizar áreas de concordância e divergência entre as classificações, evidenciando trechos reconhecidos por ambas as fontes e regiões identificadas por apenas uma delas.

Fonte: Elaboração própria

A sobreposição dos polígonos das duas bases permitiu identificar discrepâncias de

classificação, revelando áreas reconhecidas como Favelas e Comunidades Urbanas (FCU) pelo IBGE, mas não registradas como favelas no DATA.RIO, e o inverso também (Figura 2). A consolidação dessas bases possibilitou ampliar o conjunto de áreas pré-classificadas, de modo a extrair da grade de quadrados apenas os recortes de imagem que apresentavam alguma identificação prévia de favela em pelo menos uma das fontes.

Como resultado, observou-se uma redução significativa no número de imagens a serem processadas. Ao final, foram identificados 1.512 quadrados potencialmente associados à presença de áreas de favela (Figura 3).

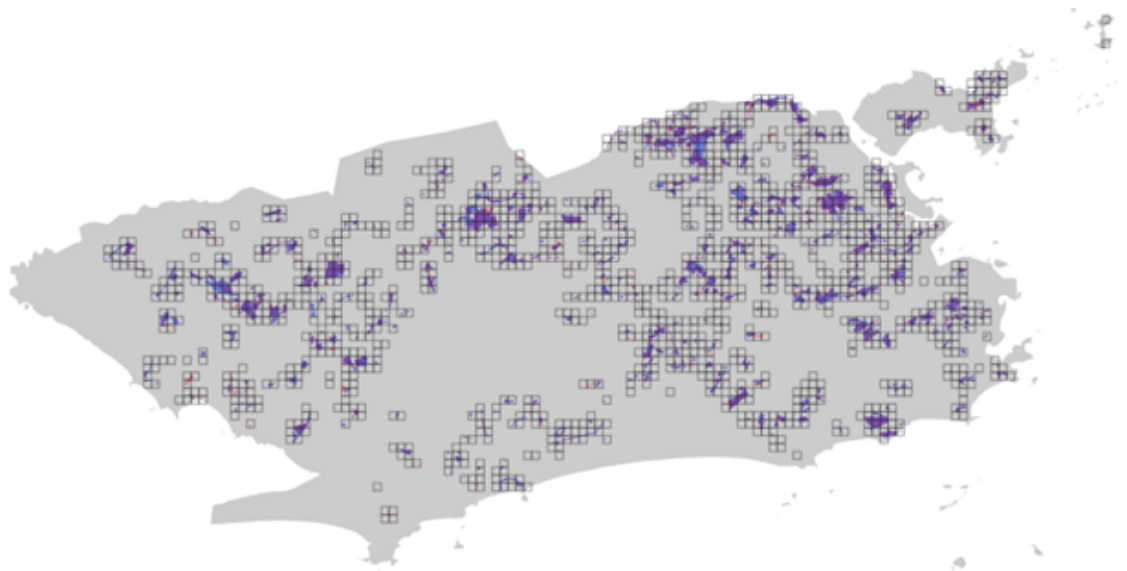


Figura 3: Sobreposição da grade de quadrados com as áreas de favelas classificadas pelo IBGE e pelo Data.Rio. A combinação dessas camadas permite identificar quais recortes da grade apresentam indícios de conter favelas dentro da cena, com base nas classificações previamente realizadas pelos órgãos.

Fonte: Elaboração própria

2.2 Amostra de Treinamento

2.2.1 Identificação das áreas de favelas

As imagens da grade espacial resultante da sobreposição das áreas do IBGE e do Data.Rio (Figura 3) foram consideradas como a população-alvo, da qual foi extraída a amostra de treinamento. Foi selecionada uma amostra estratificada com amostragem inversa, correspondente a 30% dos quadrados pertencentes às Áreas de Planejamento 1 a 4.

A Área de Planejamento 5 foi excluída nesta etapa devido ao elevado esforço necessário para a criação das máscaras de referência. Em razão dessa limitação operacional, a AP5 não foi incluída na amostra atual, sendo prevista sua incorporação em estudos futuros.

Na amostragem estratificada, a população (U) é dividida em estratos (H) — grupos distintos e exaustivos, homogêneos em relação à variável de interesse. A seleção ocorre de forma independente em cada estrato, e a amostra final resulta da união das unidades selecionadas (Silva; Bianchini; Dias, 2023).

Neste estudo, os estratos correspondem às Áreas de Planejamento, uma vez que essa divisão se baseia em critérios previamente mencionados que influenciam o padrão urbanístico. A estratificação espacial tende a aumentar significativamente a eficiência da amostragem, especialmente em regiões com características heterogêneas (Dong *et al.*, 2022).

Para a seleção dentro dos estratos, utilizou-se amostragem inversa simples, um procedimento sequencial em que, em vez de selecionar diretamente um número fixo n de unidades, avaliam-se sucessivamente as unidades da população até que n delas satisfaçam a condição de interesse (Silva; Bianchini; Dias, 2023). No presente estudo, isso correspondeu a avaliar m imagens até que n delas contivessem áreas de favelas e comunidades urbanas.

No total, foram avaliadas 358 imagens, correspondentes aos quadrados selecionados nas APs 1 a 4. Destas, 286 imagens (30% das elegíveis) compuseram a amostra final, enquanto 72 foram descartadas (Tabela 1).

Tabela 1: Quantidade de imagens da amostra nas Áreas de Planejamento

AP (H)	População (U)	Avaliadas (m)	Descartadas (r)	Amostra (n)
1	70	23	2	21
2	117	53	18	35
3	444	157	24	133
4	322	125	25	97

Fonte: Elaboração própria

A identificação das áreas de favela foi realizada por meio da análise da grade de quadrados da amostra no Google Earth Pro. Nas regiões com cobertura do Street View, procedeu-se à inspeção visual direta das edificações. Nas áreas sem essa cobertura, a avaliação foi baseada no relevo do terreno e nas construções em 3D, examinadas a partir de múltiplos ângulos e pontos de referência dentro das imagens, o que garantiu uma identificação visual consistente das áreas de interesse (Figura 4).

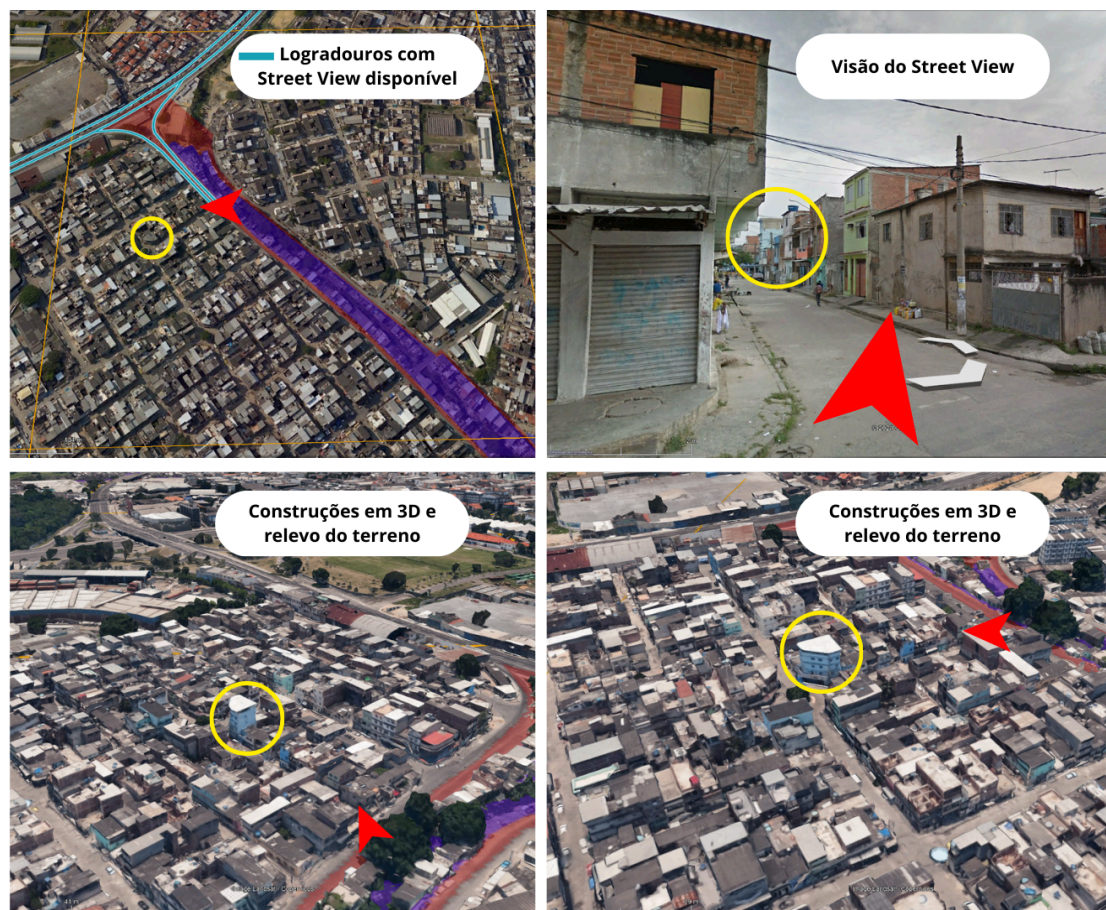


Figura 4: Visualização, no Google Earth Pro, da sobreposição da grade de quadrados com as áreas de favelas classificadas pelo IBGE e pelo Data.Rio. A análise inclui a inspeção das vias internas por meio do Street View e, nos locais onde essa ferramenta não está disponível, a observação do relevo e das construções em 3D. A fixação de elementos de referência (círculo amarelo) auxilia na localização geoespacial de cada imagem durante o processo de análise visual.

Fonte: Elaboração própria

Simultaneamente à identificação visual das favelas no Google Earth Pro, as imagens da amostra foram segmentadas manualmente no aplicativo de desenho Procreate, em um tablet, para a produção das máscaras de referência utilizadas no treinamento e teste do modelo. Nas máscaras, as áreas de favelas foram preenchidas com a cor branca, enquanto as demais áreas urbanas receberam a cor preta. Durante o processo, buscou-se contornar elementos que, embora localizados dentro das favelas, não constituem habitações, como árvores, ruas e campos de futebol, de modo a refinar o delineamento das áreas de interesse (Figura 5).

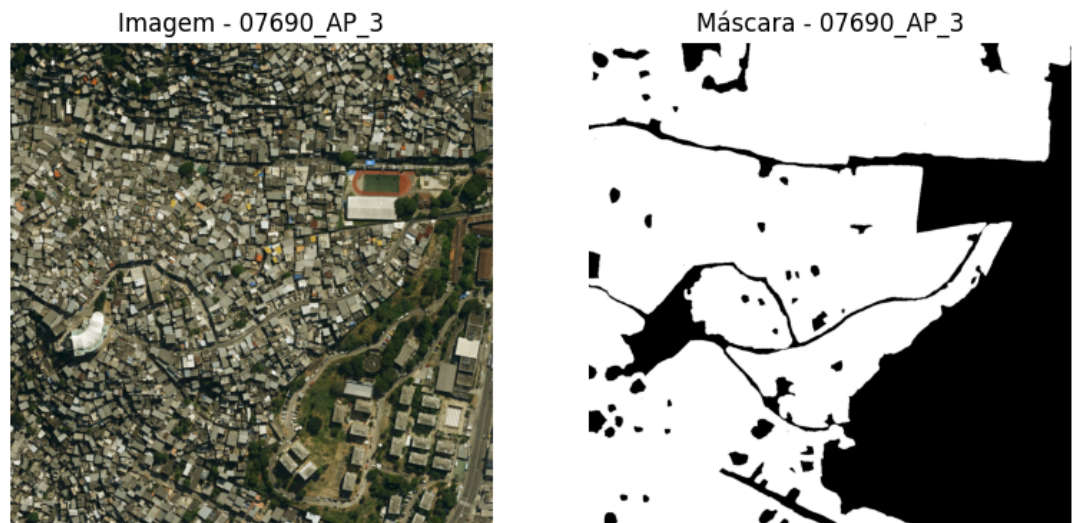


Figura 5: Imagem de satélite 7690, correspondente a parte do Complexo do Alemão (Área de Planejamento 3). No canto inferior direito, observa-se o Colégio Estadual Jornalista Tim Lopes e, um pouco acima, a Vila Olímpica Carlos Castilho. A figura apresenta também a máscara de referência produzida no estudo: as áreas em branco correspondem às regiões classificadas como favela após o processo de análise e segmentação manual da imagem, enquanto as áreas em preto representam as demais regiões urbanas não classificadas como favela.

Fonte: Elaboração própria

Ao todo, foi analisada uma área aproximada de 75 km², cujas imagens foram integralmente segmentadas de forma manual. Do total da área segmentada, 12,10% correspondem à variável de interesse, isto é, às áreas de favelas. Esse percentual é superior ao das áreas oficialmente demarcadas como favelas pelo IBGE e pelo IPP quando considerada a totalidade da extensão territorial do município do Rio de Janeiro, nas quais essas áreas representam menos de 5%.

A amostra foi dividida em dois grupos distintos: treino e teste. Do total, 80% das imagens (229) foram destinadas ao treinamento do modelo, enquanto 20% (57) compuseram o conjunto de teste (Tabela 2). As imagens foram distribuídas entre as quatro Áreas de Planejamento (Figura 6).

Tabela 2: Quantidade de imagens de treino e teste

AP	Treino	Teste
1	17	4
2	28	7
3	106	27
4	78	19

Fonte: Elaboração própria

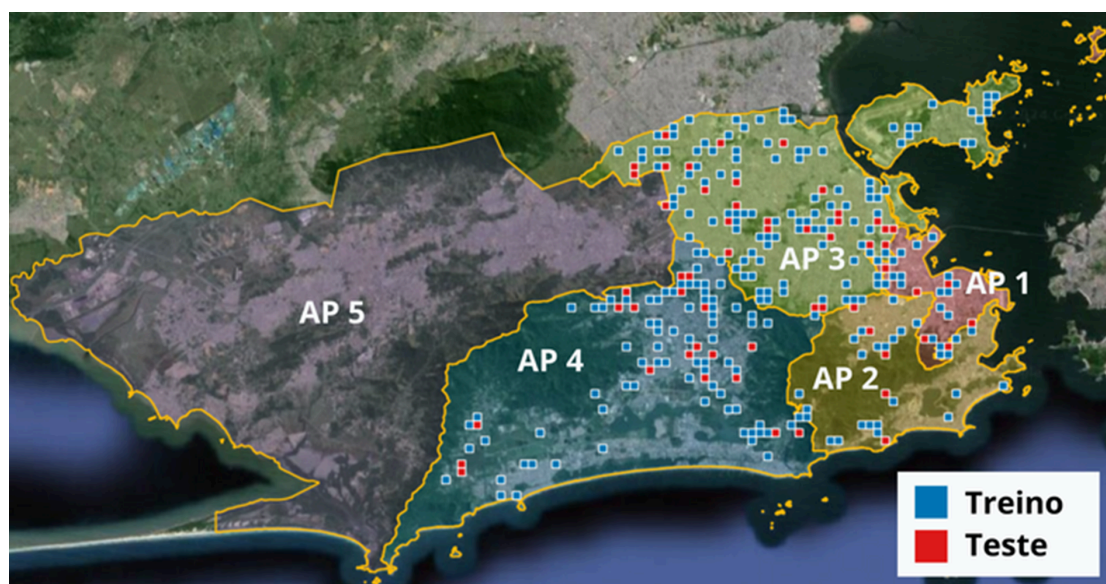


Figura 6: Distribuição espacial dos conjuntos de treino e teste pelas Áreas de Planejamento 1 a 4, correspondentes aos estratos definidos no processo de amostragem.

Fonte: Elaboração própria

2.2.2 Critérios para seleção de áreas de favelas

O descarte de parte das imagens durante o processo de amostragem inversa foi realizado por meio de inspeção visual individual, com o objetivo de garantir que apenas aquelas que apresentassem pixels correspondentes a construções características de favelas fossem incluídas na amostra.

A exclusão de imagens ocorreu pela ausência de construções típicas de favelas dentro da cena analisada. Esse fato decorre de que as delimitações espaciais utilizadas — tanto do IBGE quanto do Data.Rio — frequentemente apresentam extensão territorial superior à área efetiva da favela. Dessa forma, no processo de sobreposição dessas áreas com a grade de quadrados utilizada para definir a população, foram incluídas imagens que não continham favelas (Figura 7).

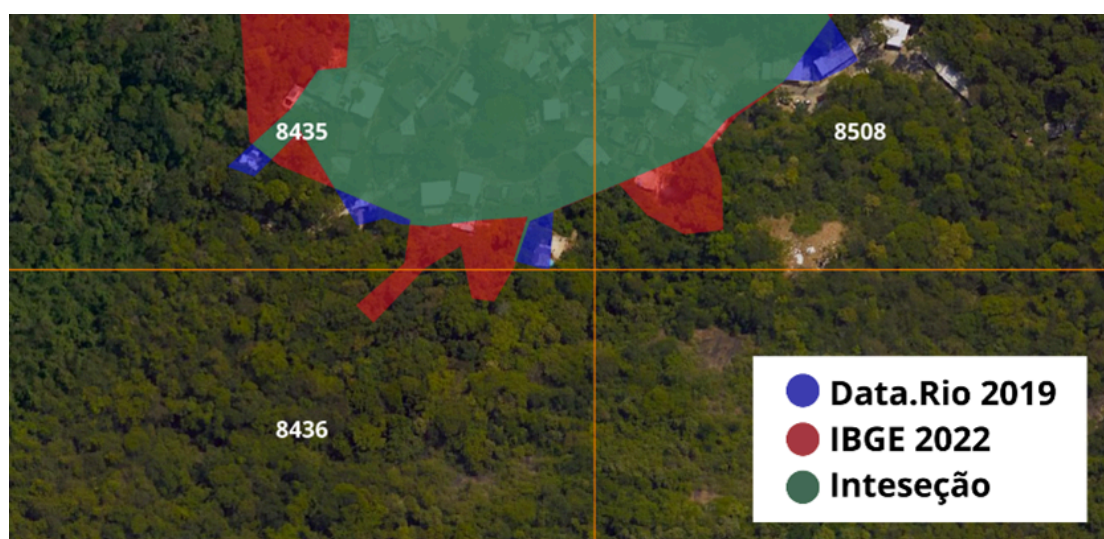


Figura 7: A imagem 8436, localizada nas proximidades do Complexo do Turano, no bairro do Rio Comprido, foi descartada. Embora a base do IBGE indique uma pequena área de favela dentro do recorte, não há construções com características compatíveis nessa porção da cena. Diferentemente, as imagens 8435 e 8508 apresentam edificações que justificam sua inclusão na análise.

Fonte: Elaboração própria

Por se tratar de uma análise baseada em imagens, a identificação das áreas de favela fundamentou-se em um conjunto de critérios visuais (Figura 8):

1. **Padrão irregular das construções:** disposição desordenada das edificações, sem indícios de planejamento urbano.
2. **Materiais construtivos precários:** uso de materiais improvisados ou de baixa durabilidade.
3. **Localização em encostas:** presença de construções em terrenos inclinados, muitas vezes sujeitos a riscos de deslizamentos.
4. **Proximidade de rodovias e ferrovias:** ocupações em faixas restritas junto a infraestruturas de transporte.
5. **Proximidade de valões:** ocorrência de edificações junto a esgoto a céu aberto ou cursos d'água canalizados de forma precária.
6. **Arruamentos estreitos:** ruas de largura reduzida, geralmente incompatíveis com a circulação de veículos de emergência ou transporte público.



Figura 8: Critérios visuais para identificação de favelas. As áreas destacadas apresentam: (1) padrão irregular de construções sem planejamento urbano; (2) materiais construtivos precários; (3) ocupação em encostas; (4) proximidade de rodovias e ferrovias em faixas restritas; (5) edificações junto a valões e esgotos a céu aberto; (6) arruamentos estreitos incompatíveis com circulação de veículos de emergência

Fonte: Elaboração própria

2.3 Aumento de Imagens para Ampliação do Conjunto de Treinamento

Com o objetivo de ampliar o conjunto de treinamento sem a necessidade de produzir novas máscaras manualmente, foi utilizada a biblioteca *Albumentations* (Buslaev, 2018), que permite aplicar transformações rápidas, eficientes e flexíveis nas imagens, como espelhamentos, rotações e distorções geométricas.

Foram aplicadas ao conjunto original (OR) de imagens (Figura 9) as seguintes transformações:

- **HorizontalFlip (HF)** – espelhamento horizontal da imagem;
- **VerticalFlip (VF)** – espelhamento vertical da imagem;
- **Transpose (TP)** – transposição da imagem, invertendo linhas e colunas.

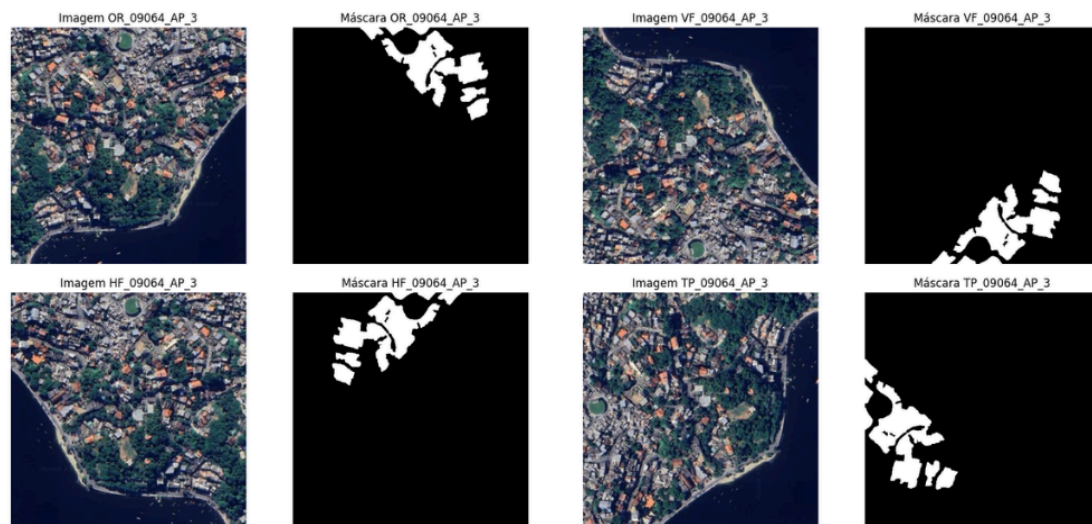


Figura 9: Imagem 09064, localizada no bairro de Cocotá, na Ilha do Governador, nas proximidades da Praia de Cocotá (Área de Planejamento 3). A imagem original foi submetida, por meio da biblioteca Albumentations, às transformações de espelhamento horizontal, espelhamento vertical e transposição de linhas e colunas.

As mesmas transformações foram aplicadas à máscara de referência correspondente, assegurando a consistência do par imagem–máscara após o aumento de dados.

Fonte: Elaboração própria

A partir dessas transformações, foram definidos dois conjuntos de dados para o treinamento do modelo:

1. **Modelo treinado sem *data augmentation*;**
2. **Modelo treinado com *data augmentation*.**

No modelo treinado com *data augmentation*, após o aumento de imagens, o número de imagens de treinamento tornou-se quatro vezes superior ao do conjunto original (sem *data augmentation*), enquanto o conjunto de teste permaneceu inalterado.

2.4 Identificação Automática das Áreas de Favelas

A identificação automática das áreas de favelas nas imagens de satélite foi realizada a partir do conjunto de imagens de treino, processado por meio de um modelo baseado na arquitetura U-Net (Ronneberger; Fischer; Brox, 2015). Esta rede neural convolucional (CNN) é projetada especificamente para tarefas de segmentação semântica, nas quais cada pixel da imagem é classificado em uma categoria específica.

A aplicação de redes neurais profundas, em particular a arquitetura U-Net, para o mapeamento de assentamentos precários a partir de imagens de satélite tem sido explorada em estudos recentes, demonstrando resultados promissores.

Lu *et al.* (2021) propuseram o GASlum-Net, uma arquitetura inovadora que combina os princípios da U-Net e da ConvNeXt em uma estrutura de duas correntes para integrar imagens RGB (do satélite Jilin-1, 5 m) e características geoespaciais derivadas do Sentinel-2. Seu modelo, treinado em 2.892 imagens de 64x64 pixels, superou significativamente as *baselines* (U-Net, ConvNeXt-UNet e FuseNet), alcançando uma melhoria de até 10,97% no IoU. O estudo destacou a eficácia do modelo na detecção de favelas de médio e grande porte (5 a >25 ha) e analisou o impacto de diferentes características e ajustes da arquitetura no desempenho.

Complementarmente, Abascal *et al.* (2022) aplicaram a U-Net para mapear padrões de privação urbana em Nairóbi, utilizando imagens de altíssima resolução (WorldView-3, 30 cm). O modelo alcançou uma acurácia de 0,92 e um IoU de 0,73

no conjunto de teste, validando sua eficácia para extração de feições urbanas. O estudo também revelou limitações na distinção de edificações individuais em áreas de alta densidade, onde o tamanho do pixel superava os espaços entre construções. Para superar isso, os autores recorreram a uma análise morfológica agregada, classificando áreas em níveis de privação (Alta, Média, Baixa) com uma acurácia geral de 0,71. A pesquisa conclui que a extração semiautomática de feições é viável e que a combinação de deep learning com análise espacial fornece insights valiosos para o planejamento urbano.

A principal característica dessa arquitetura é sua estrutura simétrica em formato de “U”, composta por duas partes complementares, o *encoder* e o *decoder* (Figura 10).

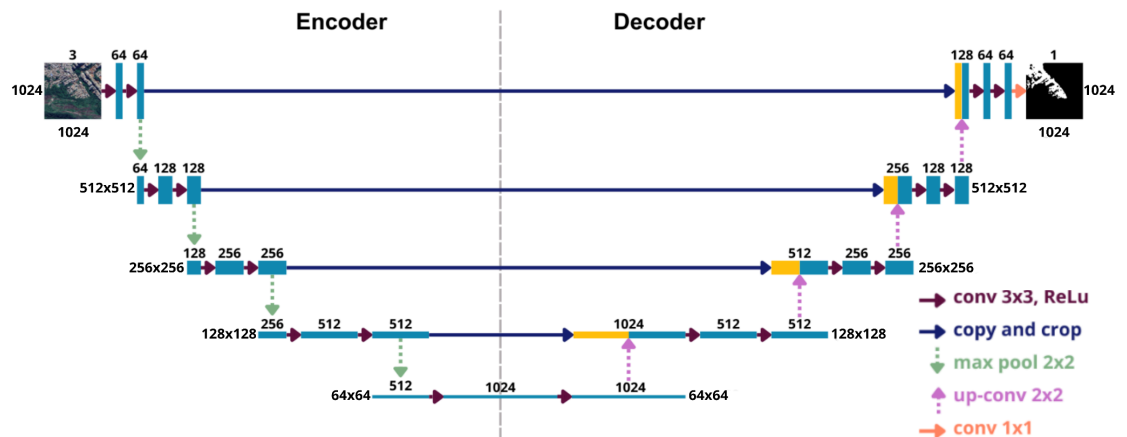


Figura 10: Ilustração da arquitetura U-Net aplicada a imagens de 1024 × 1024 pixels com três canais, apresentada em seu formato simétrico característico, composto pelos blocos de *encoder* (contração) e *decoder* (expansão).

Fonte: Elaboração própria

As caixas azuis na Figura 10 representam os mapas de características extraídos das imagens, enquanto as caixas amarelas mostram as cópias dessas características utilizadas posteriormente. As setas indicam as diferentes operações realizadas ao longo da rede.

No lado esquerdo está o *encoder*, ou caminho de contração, responsável por extrair características da imagem de entrada. À medida que a imagem passa pelas camadas convolucionais (seta vermelha) seguidas da aplicação da função de ativação ReLU (*Rectified Linear Unit*), a rede identifica padrões em diferentes níveis, aumentando o número de canais. Em seguida, as camadas de *max pooling* 2×2 (seta verde) reduzem a resolução espacial, permitindo que a rede capture informações mais gerais.

No lado direito está o *decoder*, ou caminho de expansão, responsável por recuperar a resolução das características e reconstruir o mapa de segmentação. Esse processo é realizado por meio de convoluções transpostas (seta rosa), que aumentam a resolução espacial, seguido da aplicação da função de ativação ReLU e de novas camadas convolucionais (seta vermelha), que refinam os detalhes recuperados.

Entre os dois caminhos estão as conexões de salto (seta azul), que ligam diretamente as camadas da contração às camadas correspondentes da expansão. Essas conexões concatenam os mapas de características dos dois lados, permitindo combinar detalhes finos — preservados no *encoder* — com as informações mais abstratas do *decoder*, gerando segmentações mais precisas.

Por fim, a camada de saída (seta laranja) é uma camada convolucional responsável por produzir a segmentação final, com a mesma resolução da imagem original.

Durante o treinamento do modelo, foi utilizada como função de perda a *Dice Loss*,

definida como $1 - \text{Coeficiente de Dice}$ (Figura 11). Essa escolha foi motivada pelo desbalanceamento entre as classes, uma vez que as áreas de favelas ocupam uma proporção significativamente menor da imagem em comparação às demais classes urbanas. A Dice Loss é especialmente adequada para tarefas de segmentação em cenários desbalanceados, pois penaliza de forma mais severa erros na predição da classe minoritária e favorece a maximização da sobreposição entre as máscaras preditas e de referência.

Dice Loss = 1 - Coef. Dice

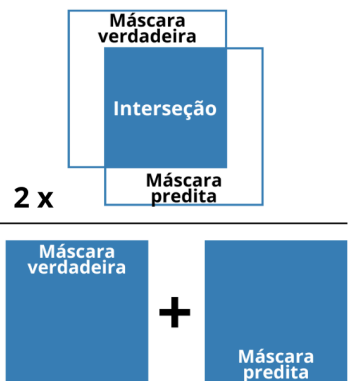
$$\text{Coeficiente de Dice} = \frac{2 \times \text{Área da interseção}}{\text{Área total}} \Leftrightarrow 2 \times \frac{\text{Área da interseção}}{\text{Área da máscara verdadeira} + \text{Área da máscara predita}}$$


Figura 11: Ilustração do conceito do *Coeficiente de Dice* e *Dice Loss*, mostrando a relação entre o dobro da área de interseção das regiões da máscara verdadeira e da máscara predita e a soma das áreas de ambas as máscaras..

Fonte: Elaboração própria

Para ambos os conjuntos de dados, o treinamento foi realizado com um número máximo de 400 épocas, utilizando *early stopping* com paciência de 50 épocas, técnica que interrompe o treinamento quando não há melhoria no desempenho em validação por um determinado número de épocas, evitando sobreajuste. Em ambos os casos, adotou-se um *batch size* de 8 e uma taxa de aprendizado de 1×10^{-4} . No Conjunto de Dados 2, o uso de *data augmentation* aumentou o número efetivo de amostras, permitindo manter essa configuração e garantindo uma comparação justa entre os modelos.

A implementação do modelo, assim como todo o pipeline de treinamento e avaliação, foi desenvolvida na linguagem Python, utilizando bibliotecas especializadas em aprendizado profundo, com destaque para TensorFlow e Keras na construção da arquitetura neural. O pré-processamento das imagens foi realizado com OpenCV (cv2) e NumPy, enquanto a avaliação quantitativa empregou métricas da biblioteca scikit-learn.

2.5 Métricas para Avaliação do Desempenho dos Modelos

As métricas de avaliação utilizadas neste estudo foram o *IoU* (*Intersection over Union*), *F1-Score* (ou Estatística F), Revocação (*Recall*) e Precisão (*Precision*). Ambas variam entre 0 e 1, sendo que valores mais próximos de 1 indicam maior precisão do modelo na tarefa de segmentação.

O *IoU*, também conhecido como índice de Jaccard, mede o grau de sobreposição entre a máscara predita e a máscara verdadeira, sendo calculado pela razão entre a área de interseção e a área de união das duas máscaras (Lakshmanan; Görner; Gillard, 2021). Quanto maior o valor do *IoU*, mais precisa é a segmentação, o que o torna especialmente relevante em tarefas que demandam alta exatidão espacial (Figura 12).

$$IoU = \frac{\text{Área da interseção}}{\text{Área da união}}$$

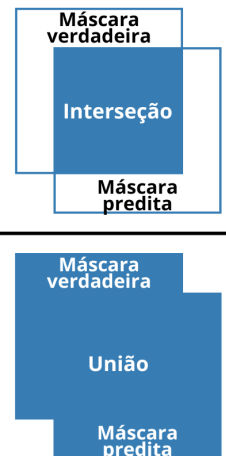


Figura 12: Ilustração do conceito de *IoU* (*Intersection over Union*), mostrando a área de interseção entre a máscara verdadeira e a máscara predita em relação à área de união entre ambas.

Fonte: Elaboração própria

Para uma compreensão mais aprofundada das demais métricas, é fundamental conhecer os conceitos associados à matriz de confusão (Figura 13): Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN) (Provost; Fawcett, 2016).

	Valor predito: Verdadeiro	Valor predito: Verdadeiro
Valor real: Verdadeiro	Verdadeiro Positivo (VP) <small>Predição correta</small>	Falso Negativo (FN) <small>Erro do Tipo II</small>
Valor real: Falso	Falso Positivo (FP) <small>Erro do Tipo I</small>	Verdadeiro Negativo (VN) <small>Predição correta</small>

Figura 13: Matriz de confusão, que compara os valores preditos do modelo com os valores reais, organizando os resultados em quatro categorias: Verdadeiros Positivos (VP) e Verdadeiros Negativos (VN) representam acertos; Falsos Positivos (FP) e Falsos Negativos (FN) representam os erros.

Fonte: Elaboração própria

- **Verdadeiro Positivo (VP):** Caso em que o modelo prevê a classe positiva e essa previsão corresponde à classe positiva real. Ou seja, o modelo identifica corretamente a presença da classe de interesse.
- **Verdadeiro Negativo (VN):** Situação na qual tanto a previsão do modelo quanto a realidade pertencem à classe negativa. Dessa forma, o modelo reconhece adequadamente a ausência da classe de interesse.
- **Falso Positivo (FP):** Ocorre quando o modelo prevê a classe positiva, mas a classe real é negativa. Em outras palavras, o modelo sinaliza erroneamente a presença da classe de interesse.
- **Falso Negativo (FN):** Caso em que o modelo prevê a classe negativa, mas a classe real é positiva. Isso significa que o modelo falha ao não identificar a presença da classe de interesse.

Precisão e Revocação são métricas complementares amplamente utilizadas na avaliação de modelos de classificação, especialmente em contextos de desbalanceamento de classes (Figura 14). Enquanto a Precisão está associada à confiabilidade das previsões positivas realizadas pelo modelo, a Revocação reflete sua capacidade de identificar corretamente os casos positivos existentes (Provost; Fawcett, 2016):

- **Precisão:** avalia a proporção de previsões positivas que são efetivamente corretas, sendo calculada pela razão entre os verdadeiros positivos (VP) e a soma dos verdadeiros positivos e dos falsos positivos (VP + FP).
- **Revocação:** mede a capacidade do modelo de identificar todos os casos positivos reais, correspondendo à razão entre os verdadeiros positivos (VP) e a soma dos verdadeiros positivos e dos falsos negativos (VP + FN).

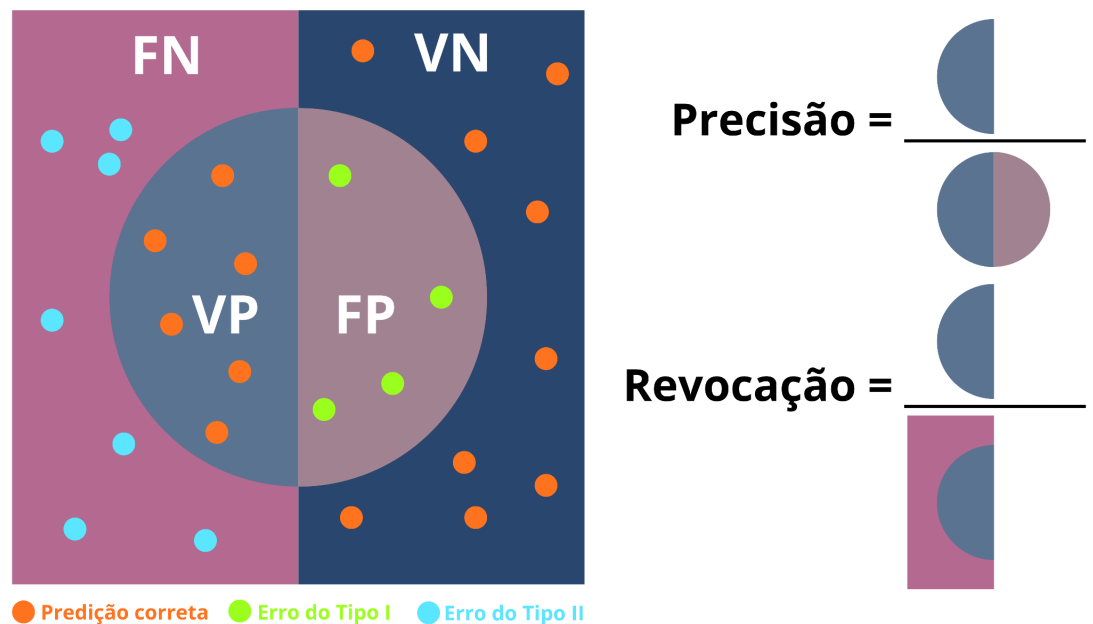


Figura 14: Ilustração do conceito de Precisão e Revocação. A figura ilustra como falsos positivos reduzem a Precisão e falsos negativos reduzem a Revocação, enfatizando que a ênfase em uma métrica impacta a outra.

Fonte: Elaboração própria

O *F1-Score*, por sua vez, integra as métricas de Precisão e Revocação em uma única medida, permitindo avaliar simultaneamente a confiabilidade das previsões positivas e a capacidade do modelo de identificar corretamente todos os casos positivos (Provost; Fawcett, 2016).

Essa métrica é calculada por meio da média harmônica entre Precisão e Revocação, o que penaliza valores extremos e resulta em uma avaliação mais equilibrada e representativa do desempenho do modelo, especialmente em cenários com desbalanceamento de classes.

$$F1 = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (1)$$

3 Resultados

Após o treinamento dos modelos com e sem data augmentation, os menores valores da função de perda foram atingidos na época 52 para o modelo sem aumento de dados e na época 71 para o modelo com aumento de dados. Em ambos os casos, não houve melhoria nas 50 épocas subsequentes, o que acionou o critério de *early stopping*.

A análise global dos resultados, considerando a média das quatro áreas de planejamento, indica que o modelo treinado com *data augmentation* apresenta desempenho superior em três das quatro métricas avaliadas. Obtendo maiores valores médios de *IoU*, *Precisão* e *F1-Score*, evidenciando uma melhora na qualidade geral da segmentação e no equilíbrio entre erros de falso positivo e falso negativo.

Por outro lado, o modelo sem *data augmentation* apresentou maior valor médio de Revocação, indicando maior capacidade de identificar corretamente as regiões de interesse, ainda que à custa de uma maior incidência de falsos positivos. Esse comportamento reforça a existência de um *trade-off* entre *Precisão* e Revocação, no qual o aumento de uma métrica tende a impactar negativamente a outra.

De forma geral, os resultados sugerem que a utilização de *data augmentation* contribui para tornar o modelo mais equilibrado e robusto, com ganhos consistentes em métricas agregadas como *IoU* e *F1-Score*, que são particularmente relevantes para a avaliação global do desempenho em tarefas de segmentação (Tabela 3).

Tabela 3. Desempenho médio das métricas de avaliação da U-Net treinada sem e com *data augmentation*. Os valores superiores entre os dois modelos por Área de Planejamento e Geral estão destacados em azul.

AP	Modelo sem <i>data augmentation</i>				Modelo com <i>data augmentation</i>			
	IoU	Prec	Revoc	F1	IoU	Prec	Revoc	F1
1	0,7037	0,5101	0,7850	0,5887	0,7053	0,5328	0,7031	0,5828
2	0,7491	0,6953	0,7962	0,7287	0,7576	0,7419	0,7659	0,7344
3	0,6334	0,5188	0,7785	0,5628	0,6847	0,6034	0,7548	0,6256
4	0,7283	0,5188	0,7785	0,6791	0,7441	0,6902	0,7282	0,6955
Geral	0,6842	0,5189	0,7751	0,6238	0,7149	0,6444	0,7437	0,6592

Fonte: Elaboração própria

A análise desagregada por Área de Planejamento (AP) evidencia que os efeitos do *data augmentation* variam conforme o contexto espacial, embora se observe um padrão geral de melhora em métricas agregadas. Nas APs 2, 3 e 4, o modelo com *data augmentation* apresentou desempenho superior ou equivalente ao modelo sem aumento de dados nas métricas de *IoU*, *Precisão* e *F1-Score*, indicando ganhos consistentes na qualidade da segmentação e no equilíbrio entre erros de classificação.

Na AP 3, em particular, a vantagem do modelo com *data augmentation* é mais pronunciada, com aumentos expressivos em *IoU*, *Precisão* e *F1-Score*, além de uma Revocação elevada. Resultados semelhantes, ainda que menos acentuados, são observados nas APs 2 e 4, onde o modelo com *data augmentation* mantém desempenho globalmente superior.

A AP 1 apresenta um comportamento distinto. Embora o modelo com *data augmentation* obtenha *IoU* ligeiramente superior (0,7053 contra 0,7037) e maior *Precisão* (0,5328 contra 0,5101), o modelo sem *augmentation* apresenta Revocação consideravelmente mais elevada (0,7850 contra 0,7031), resultando em um *F1-Score* marginalmente superior. Esse resultado indica um *trade-off* mais acentuado entre *Precisão* e Revocação nessa área específica, possivelmente associado às características particulares da região e à limitada quantidade de amostras disponíveis.

Em síntese, a análise por AP confirma que o *data augmentation* tende a produzir modelos mais equilibrados e com melhor desempenho global, especialmente em termos de *IoU* e *F1-Score*, ainda que seu impacto possa variar localmente em função da distribuição e da representatividade dos dados de treinamento.

O *boxplot* apresenta a distribuição das métricas para as 57 imagens do conjunto de teste, estratificadas por Área de Planejamento, observam-se padrões importantes (Figura 15). O modelo treinado com *data augmentation* demonstra menor dispersão nos valores de *IoU* e *F1-Score*, indicando uma performance mais consistente e previsível nessas métricas. Em contrapartida, o modelo sem *data augmentation* apresenta valores medianos superiores na métrica de Revocação. Quanto à Precisão, o modelo com *augmentation* mostra uma distribuição levemente deslocada para cima, com primeiro, terceiro quartis e mediana mais elevados, sugerindo um desempenho superior.

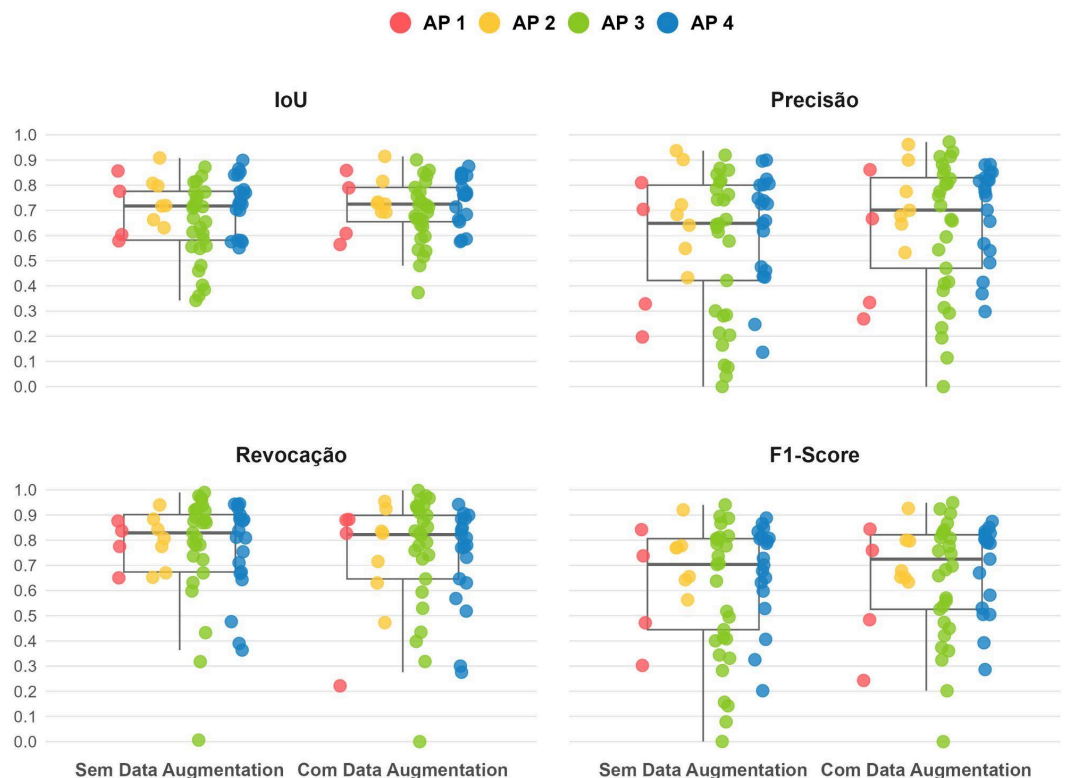


Figura 15: Distribuição das métricas de avaliação (*IoU*, Precisão, Revocação e *F1-Score*) da U-Net treinada sem e com *data augmentation*, estratificada por Área de Planejamento (AP). A análise revela que o modelo com aumento de dados apresenta menor dispersão nas métricas de *IoU*, Precisão e *F1-Score*, enquanto o modelo sem *augmentation* quartis superiores em Revocação.

Fonte: Elaboração própria

Para avaliar objetivamente se o ganho em desempenho observado era estatisticamente significativo, foram conduzidos testes de hipóteses comparando os valores de *IoU* e *F1-Score* obtidos pelo modelo com *data augmentation* contra os do modelo sem *data augmentation*. Como os dados das 57 imagens do conjunto de teste representam medições pareadas (cada imagem avaliada por ambos os modelos), testes estatísticos adequados para amostras pareadas foram selecionados.

Inicialmente, o teste de *Shapiro–Wilk* aplicado às diferenças individuais dos valores de *IoU* para cada imagem rejeitou a hipótese de normalidade (valor $p = 0,0019$), violando assim um pressuposto fundamental do teste *t* pareado. Diante disso, optou-se pelo teste de *Wilcoxon* para amostras pareadas, um método não paramétrico robusto a desvios da normalidade, que avalia se as diferenças entre os pares apresentam um deslocamento sistemático em relação a zero.

O teste foi configurado com hipótese alternativa (H_1) unilateral de que os valores de *IoU* do modelo com *data augmentation* (M_2) tendem a ser maiores do que os do modelo sem *data augmentation* (M_1), conforme definido abaixo:

- H_0 : As distribuições dos valores de *IoU* de M_1 e M_2 são iguais. Ou seja, não há evidência de que os valores de *IoU* de M_2 sejam sistematicamente maiores que os de M_1 .
- H_1 : Os valores de *IoU* de M_2 são sistematicamente maiores que os de M_1 .

Considerando um nível de significância (α) de 0,05, o valor p de 0,0006 permitiu rejeitar a hipótese nula (H_0), fornecendo evidência estatística de que os valores de *IoU* do modelo treinado com *data augmentation* são superiores. A mediana das diferenças foi positiva (+0,0183), corroborando essa conclusão. Adicionalmente, em 71,9% das imagens (41 de 57), o modelo com *data augmentation* superou o modelo sem *augmentation*, reforçando a consistência prática da vantagem observada.

Para o *F1-Score*, procedimento semelhante foi adotado. O teste de *Shapiro–Wilk* também rejeitou a normalidade das diferenças (valor p = 0,0002), justificando novamente o uso do teste de *Wilcoxon* não paramétrico. Aplicou-se o mesmo esquema de hipóteses:

- H_0 : As distribuições dos valores de *F1-Score* de M_1 e M_2 são iguais.
- H_1 : Os valores de *F1-Score* de M_2 são sistematicamente maiores que os de M_1 .

O resultado, considerando também um nível de significância (α) de 0,05, com valor p = 0,0060, a hipótese nula foi rejeitada. A superioridade do modelo com *data augmentation* também se manifestou na diferença das medianas do *F1-Score* (+0,0122 pontos). O modelo M_2 superou M_1 em 66,67% das imagens (38 de 57), apresentando desempenho inferior em apenas 33,33% dos casos (19 de 57).

Com o objetivo de visualizar e compreender o comportamento das previsões na identificação de favelas em imagens de satélite, são apresentados a seguir resultados visuais referentes a duas amostras de cada uma das quatro Áreas de Planejamento (AP): um exemplo de predição bem-sucedida e outro em que o desempenho foi inferior. Cada imagem é acompanhada de suas respectivas métricas de *IoU* e *F1-Score*. Para facilitar a análise comparativa, as áreas preditas como favela foram sobrepostas às imagens originais com aplicação de transparência, permitindo visualizar claramente as regiões urbanas classificadas pelo modelo U-Net e, ao mesmo tempo, preservar a forma original para comparação direta com as máscaras de referência.

Na imagem 08135, correspondente à Área de Planejamento 1, observa-se no canto superior direito um trecho da favela Parque Horácio Cardoso Franco, localizada em Benfica, na Rua Couto de Magalhães (Figura 16). Ambos os modelos apresentaram desempenho satisfatório na segmentação da área de interesse; contudo, o modelo treinado com *data augmentation* obteve resultados superiores, refletidos em maiores valores de *IoU* e *F1-Score*. Observa-se que o modelo sem *data augmentation* classificou erroneamente uma pequena região no canto superior direito da imagem, enquanto os erros do modelo com aumento de dados mostraram-se mais localizados e restritos às proximidades da favela, indicando uma segmentação mais precisa e espacialmente coerente.

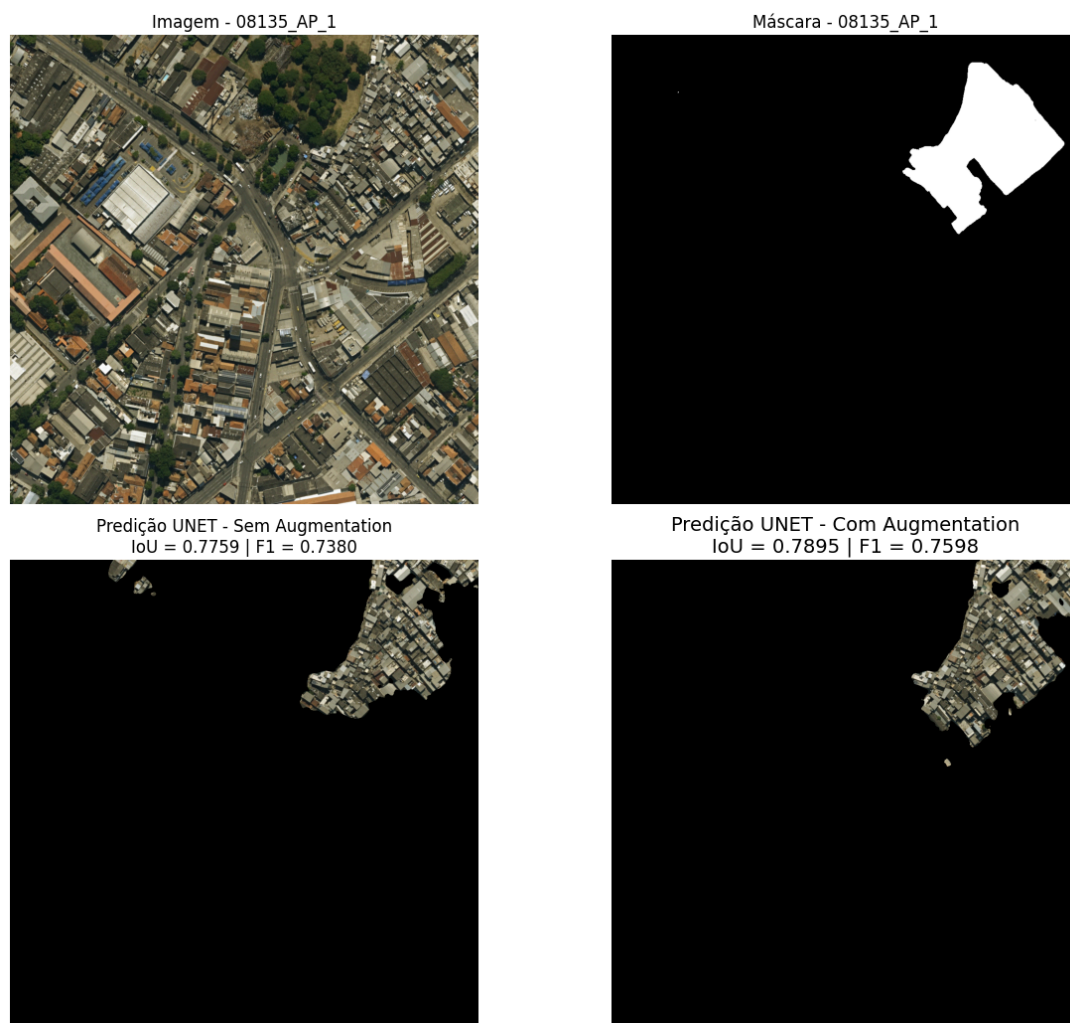


Figura 16: Comparação visual das predições para a imagem 08135 (Área de Planejamento 1). Da esquerda para a direita: (a) imagem de satélite original; (b) máscara de referência; (c) predição do modelo sem *data augmentation*; (d) predição do modelo com *data augmentation*. A cena inclui, no canto superior direito, um segmento da favela Parque Horácio Cardoso Franco, em Benfica (Rua Couto de Magalhães).

Fonte: Elaboração própria

A imagem 08430, correspondente à Área de Planejamento 1, contém, em sua porção inferior, uma pequena favela registrada exclusivamente na base de dados do Instituto Pereira Passos, localizada na esquina da Rua José Eugênio com a Rua Francisco Eugênio, no bairro de São Cristóvão (Figura 17). Em função da reduzida dimensão da área de interesse, ambos os modelos apresentaram baixa sobreposição com a região de favela, ainda que com valores de *IoU* superiores a 0,5. O modelo sem *data augmentation* obteve desempenho superior tanto em *IoU* quanto em *F1-Score*; entretanto, apresentou uma área significativamente maior classificada erroneamente como favela quando comparado ao modelo treinado com aumento de dados, evidenciando um *trade-off* entre sobreposição global e controle de falsos positivos.

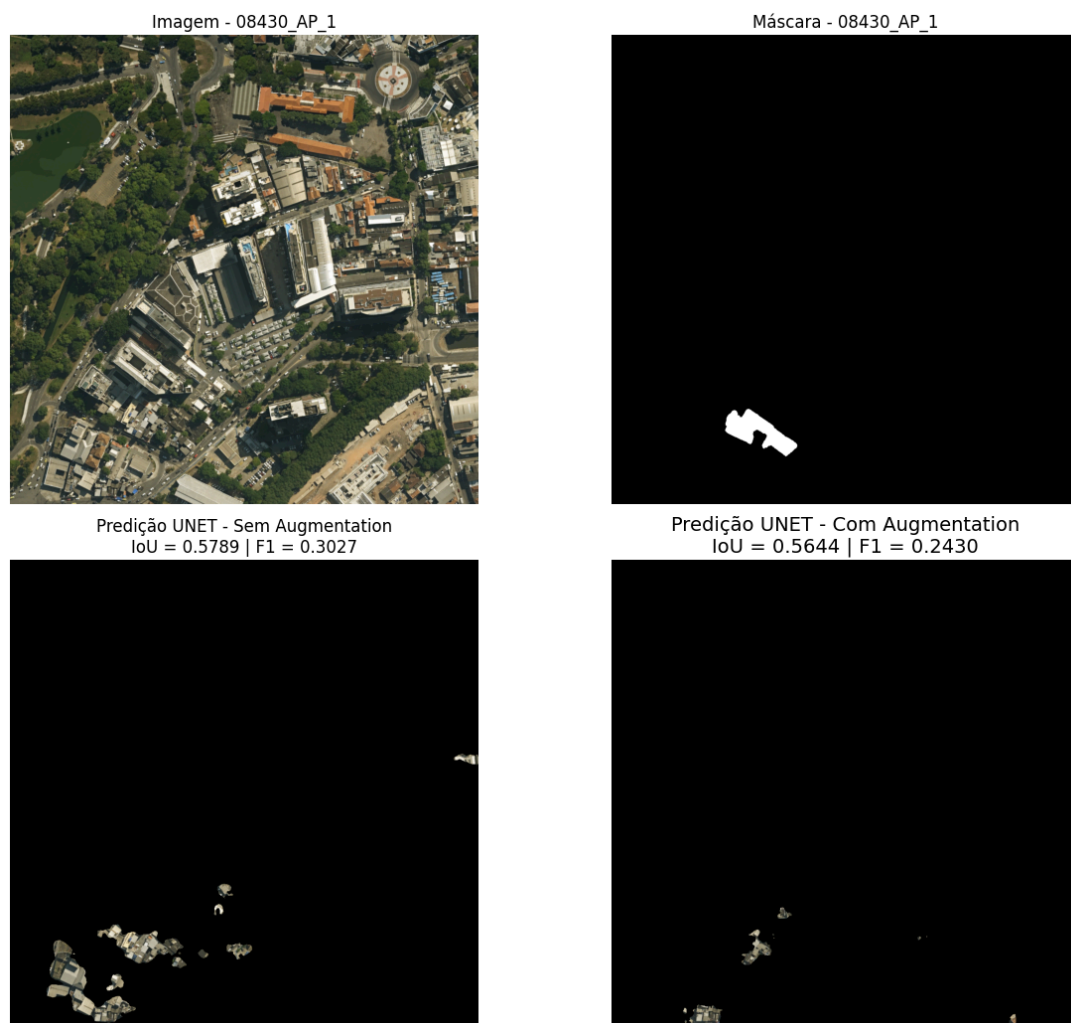


Figura 17: Comparação visual das predições para a imagem 08430 (Área de Planejamento 1). Da esquerda para a direita: (a) imagem de satélite original; (b) máscara de referência; (c) predição do modelo sem *data augmentation*; (d) predição do modelo com *data augmentation*. A cena contém, na porção inferior, uma pequena favela catalogada apenas pelo Instituto Pereira Passos, situada na esquina da Rua José Eugênio com a Rua Francisco Eugênio, em São Cristóvão.
Fonte: Elaboração própria

Na imagem 07848, referente à Área de Planejamento 2, observam-se trechos do Parque Vila Isabel, separados na cena pelo Parque Recanto do Trovador e pela Vila Olímpica Artur da Távola, no bairro de Vila Isabel, nas proximidades da Rua Visconde de Santa Isabel com a Rua Barão do Bom Retiro (Figura 18). Ambos os modelos apresentaram desempenho elevado, com valores de *IoU* e *F1-Score* superiores a 0,9. Ainda assim, o modelo treinado com *data augmentation* demonstrou ligeira superioridade, ao delimitar com maior precisão os contornos da favela e discriminar adequadamente elementos internos que não correspondem a domicílios precários. Nota-se que pequenos trechos na porção inferior das predições de ambos os modelos foram classificados erroneamente como favela; tais regiões correspondem, na realidade, a áreas de terraço em edificações residenciais de pequeno porte.

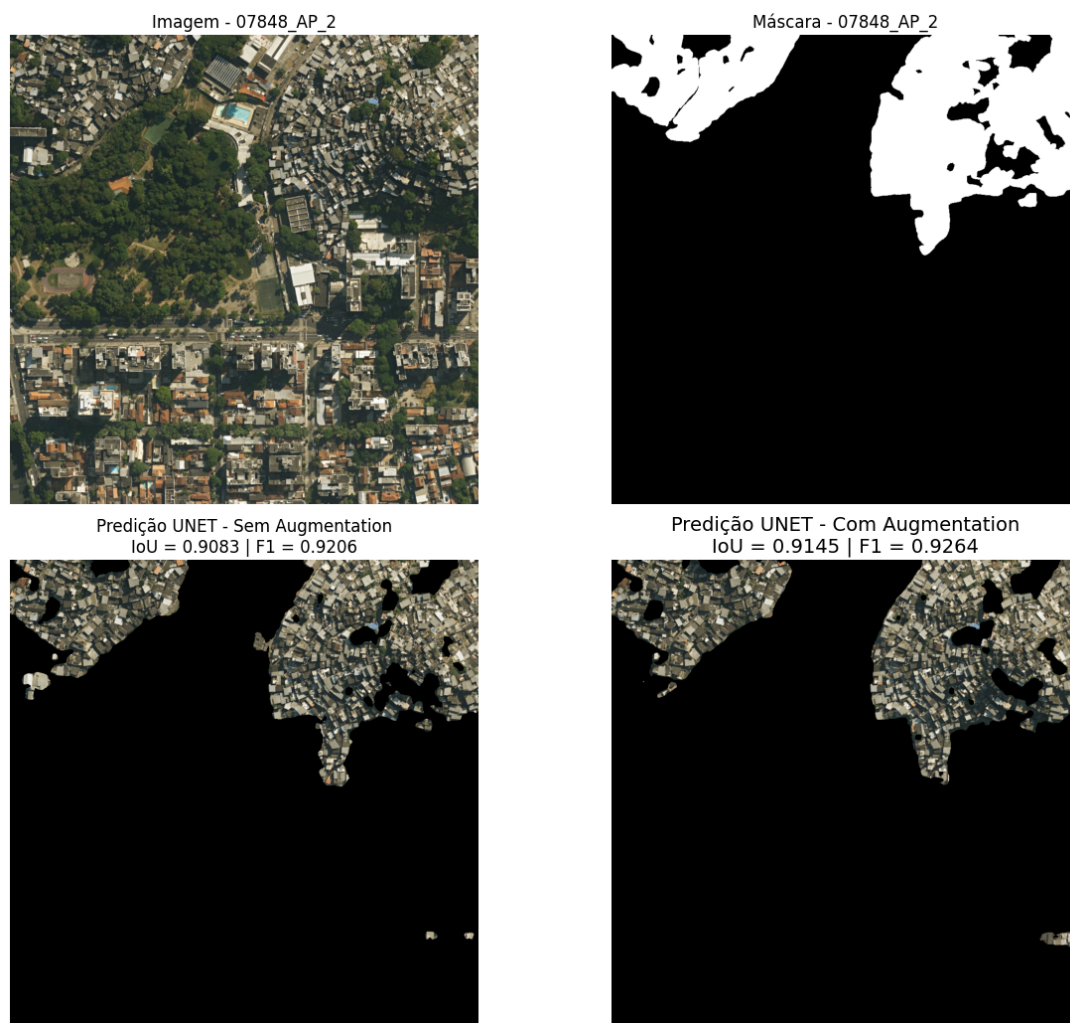


Figura 18: Comparação visual das predições para a imagem 07848 (Área de Planejamento 2). Da esquerda para a direita: (a) imagem de satélite original; (b) máscara de referência; (c) predição do modelo sem *data augmentation*; (d) predição do modelo com *data augmentation*. A cena exibe trechos do Parque Vila Isabel, separados pelo Parque Recanto do Trovador e pela Vila Olímpica Artur da Távola, no bairro de Vila Isabel, próximo ao cruzamento das ruas Visconde de Santa Isabel e Barão do Bom Retiro.

Fonte: Elaboração própria

Na imagem 08157, referente à Área de Planejamento 2, observa-se a favela Vidigal, localizada no bairro de mesmo nome, próximo à Avenida Niemeyer (Figura 19). Nota-se que ambos os modelos classificaram erroneamente uma extensa área como favela, a qual não corresponde à máscara de referência nem está registrada como tal nas bases do IPP ou do IBGE. Esta região apresenta um padrão misto de ocupação: embora contenha construções que claramente não se enquadram na classificação de favela, outras possuem características visuais ambíguas que poderiam suscitar questionamentos. Adicionalmente, a morfologia acidentada do terreno, com forte inclinação, contribui para uma certa irregularidade nas formas dos lotes e das edificações, o que pode ter influenciado os erros de classificação.

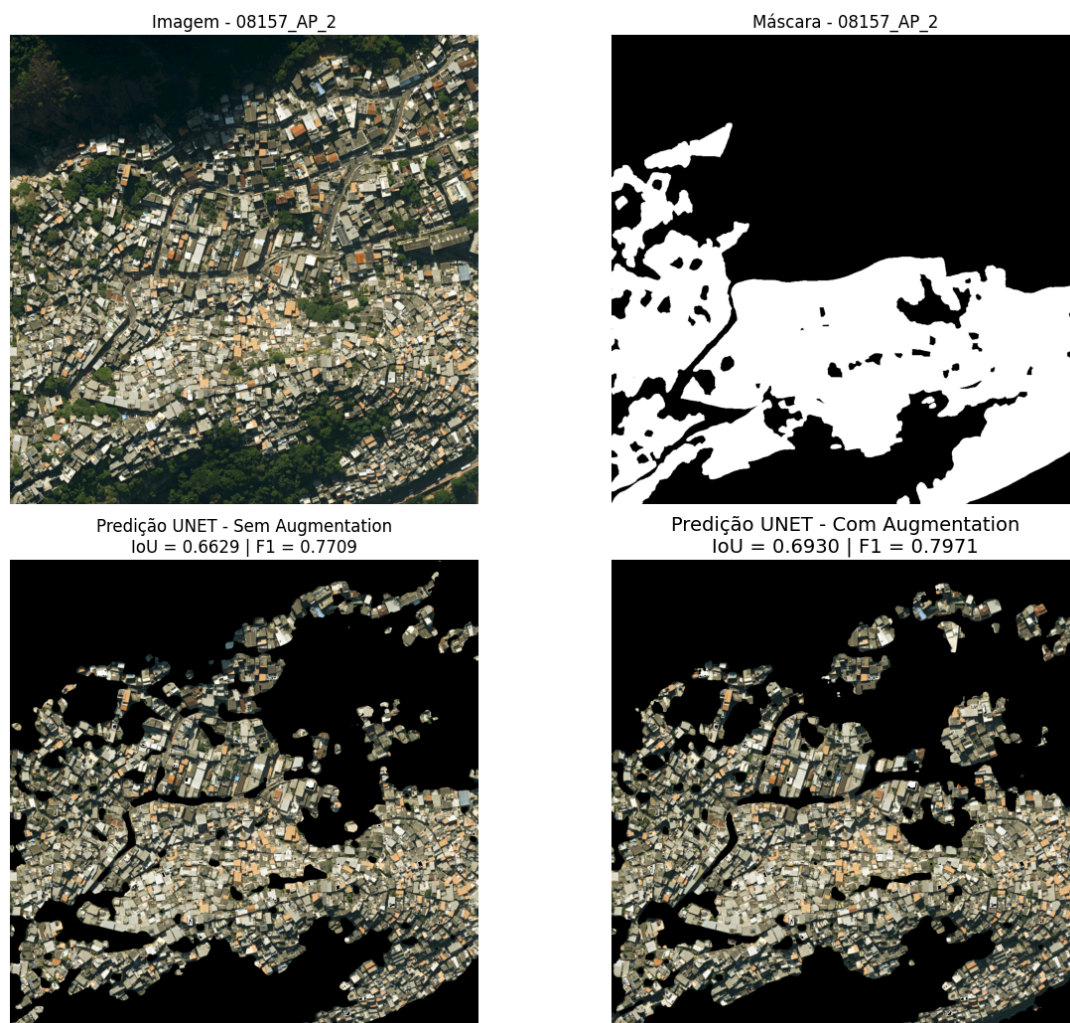


Figura 19: Comparação visual das predições para a imagem 08157 (Área de Planejamento 2). Da esquerda para a direita: (a) imagem de satélite original; (b) máscara de referência; (c) predição do modelo sem *data augmentation*; (d) predição do modelo com *data augmentation*. A cena mostra a favela Vidigal, no bairro homônimo, nas proximidades da Avenida Niemeyer.

Fonte: Elaboração própria

A imagem 07691 mostra a favela Nova Brasília, localizada no Complexo do Alemão na Área de Planejamento 3, nas proximidades do cruzamento da Estrada de Itararé com a Avenida Itaóca (Figura 20). A predição gerada pelo modelo com *data augmentation* apresentou desempenho superior, caracterizado por uma menor ocorrência de falsos positivos (áreas não favela erroneamente classificadas) em comparação ao modelo sem aumento de dados. As áreas incorretamente segmentadas como favela por ambos os modelos correspondem, em sua maioria, a condomínios residenciais.

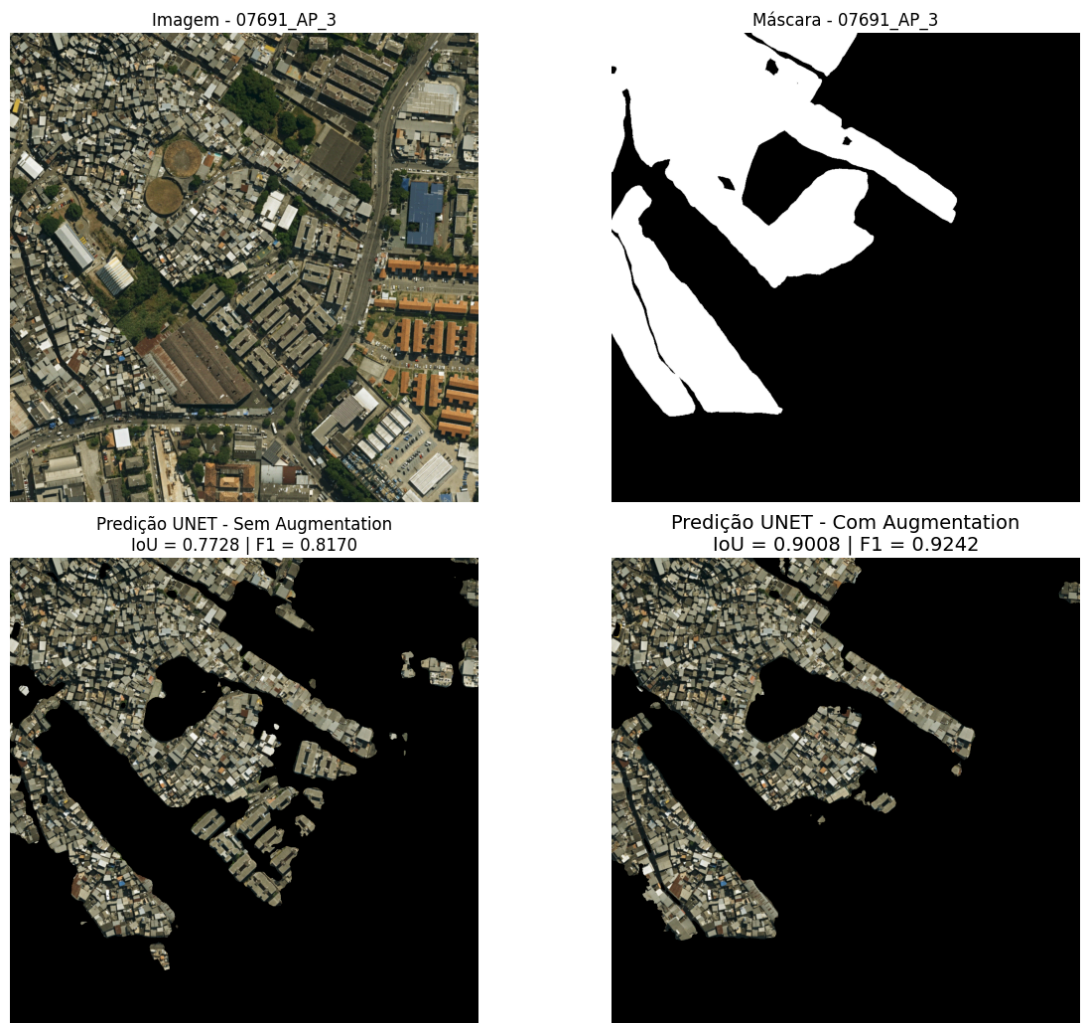


Figura 20: Comparação visual das predições para a imagem 07691 (Área de Planejamento 3). Da esquerda para a direita: (a) imagem de satélite original; (b) máscara de referência; (c) predição do modelo sem *data augmentation*; (d) predição do modelo com *data augmentation*. A cena exibe a favela Nova Brasília, no Complexo do Alemão, próxima ao encontro da Estrada de Itararé com a Avenida Itaóca.

Fonte: Elaboração própria

A imagem 07170, da Área de Planejamento 3, exibe a pequena favela Castelo de Lucas, registrada apenas na base de dados do IPP e localizada no bairro Parada de Lucas, no encontro da Rua Cordovil com a Rua Amadeu Amaral (Figura 21). Em uma área reduzida como está, ambos os modelos tiveram uma baixa sobreposição com *IoU* inferior a 0,5. O modelo sem *data augmentation* ainda conseguiu uma sobreposição com a área de favela evidenciado pelo *F1-Score* de 0,0006, no entanto classificou erroneamente como favela diversas áreas. Já o modelo com *data augmentation* classificou erroneamente uma menor área como favela, mas teve um *F1-Score* zero.

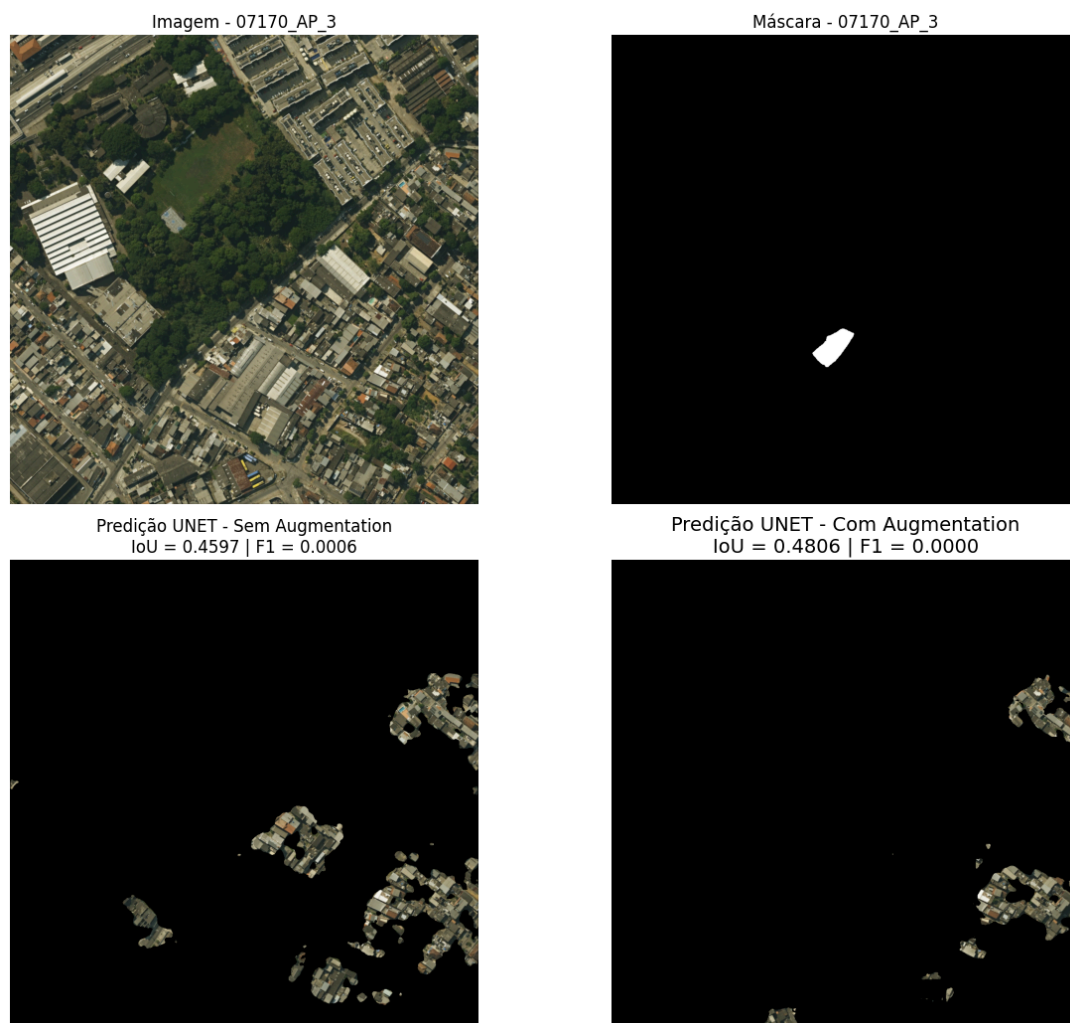


Figura 21: Comparação visual das predições para a imagem 07170 (Área de Planejamento 3). Da esquerda para a direita: (a) imagem de satélite original; (b) máscara de referência; (c) predição do modelo sem *data augmentation*; (d) predição do modelo com *data augmentation*. A cena mostra a pequena favela Castelo de Lucas, registrada exclusivamente pelo IPP, situada no bairro Parada de Lucas, na confluência das ruas Cordovil e Amadeu Amaral.

Fonte: Elaboração própria

A imagem 05804, da Área de Planejamento 4, mostra a favela Comunidade Corumau, localizada no bairro da Taquara, defronte à Estrada Curumau (Figura 22). Ambos os modelos alcançaram valores elevados de *IoU* e *F1-Score*; entretanto, o modelo treinado com *data augmentation* apresentou desempenho superior. Do ponto de vista qualitativo, observa-se que esse modelo delimitou com maior fidelidade os contornos da favela e gerou uma quantidade menor de falsos positivos — isto é, áreas não pertencentes à favela classificadas incorretamente — quando comparado ao modelo sem aumento de dados. As principais regiões erroneamente segmentadas correspondem a terraços de sobrados e de edificações residenciais de pequeno porte, cujas características espectrais e texturais podem se assemelhar às de áreas de favela em imagens de satélite, dificultando a distinção por ambos os modelos.

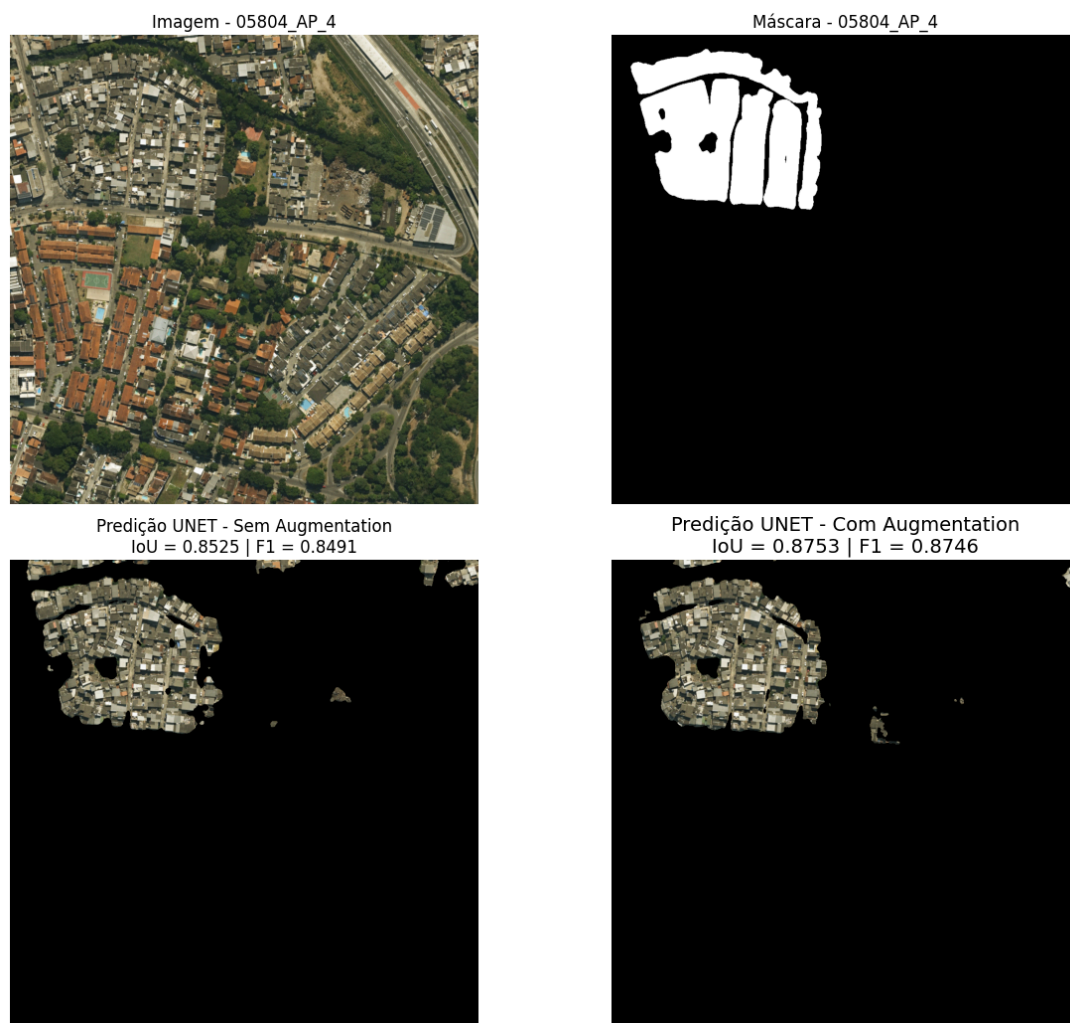


Figura 22: Comparação visual das predições para a imagem 05804 (Área de Planejamento 4). Da esquerda para a direita: (a) imagem de satélite original; (b) máscara de referência; (c) predição do modelo *sem data augmentation*; (d) predição do modelo com *data augmentation*. A cena exibe a favela Comunidade Corumau, na Taquara, de frente para a Estrada Curumau.

Fonte: Elaboração própria

A imagem 05729, da Área de Planejamento 4, mostra um pequeno aglomerado residencial classificado como favela apenas na base do IPP, situado defronte à Estrada da Boiúna, no bairro da Taquara, no canto inferior esquerdo da cena (Figura 23). Ambos os modelos conseguiram identificar parcialmente esse conjunto de edificações, atribuindo-lhe corretamente a classe de favela. No entanto, também foram observados falsos positivos em outras regiões que não constam como favela nem na máscara de referência nem nas bases oficiais. A área segmentada no canto superior esquerdo da imagem, nas proximidades da Rua Pereira, apresenta características morfológicas e espectrais ambíguas, o que pode justificar a divergência de classificação e suscitar questionamentos quanto à sua real categorização. Por outro lado, as edificações identificadas no canto inferior direito apenas pelo modelo *sem data augmentation* correspondem, na realidade, a um pequeno condomínio situado na Estrada Curumau, próximo ao Túnel Senador Nelson Carneiro, configurando um erro inequívoco de classificação.

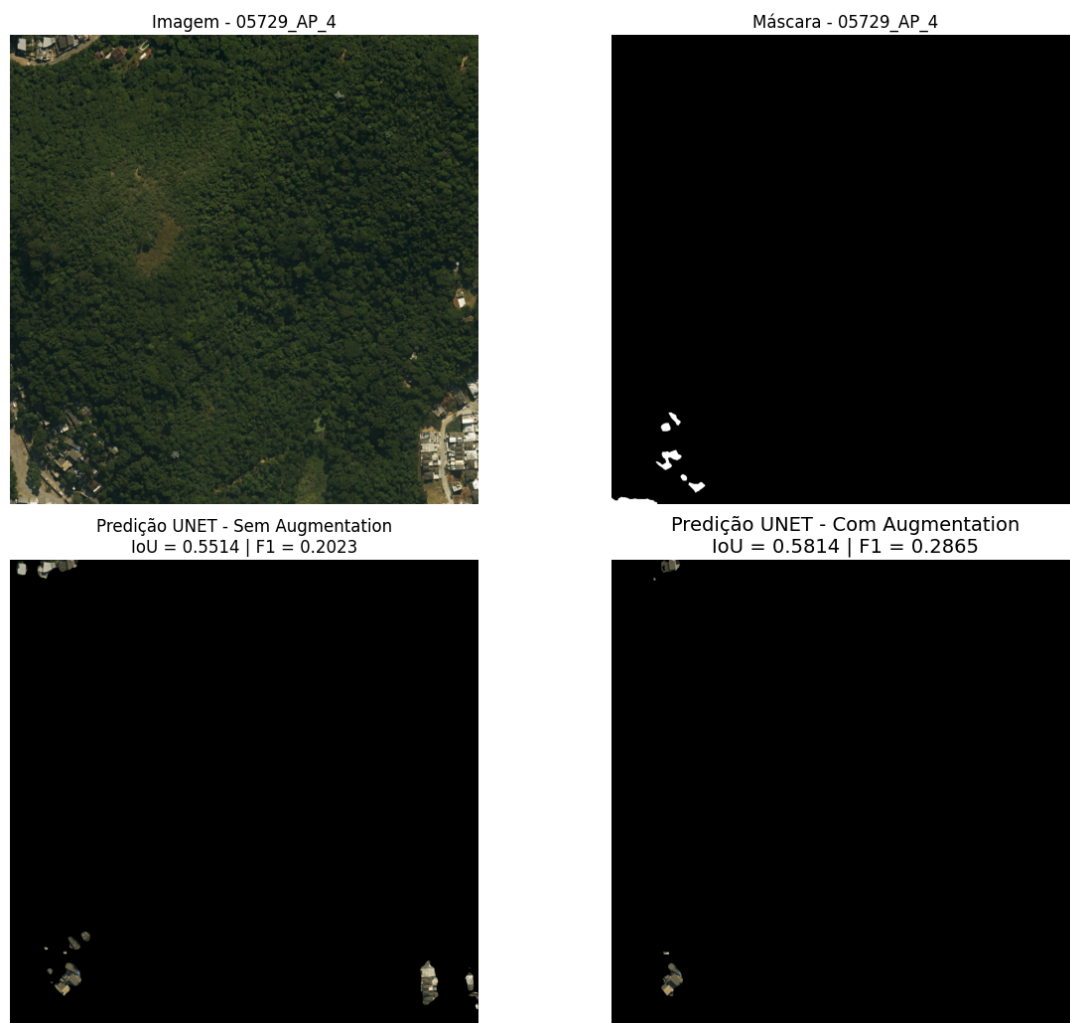


Figura 23: Comparação visual das predições para a imagem 05729 (Área de Planejamento 4). Da esquerda para a direita: (a) imagem de satélite original; (b) máscara de referência; (c) predição do modelo sem *data augmentation*; (d) predição do modelo com *data augmentation*. A cena exibe, no canto inferior esquerdo, um pequeno aglomerado residencial catalogado como favela apenas pelo IPP, defronte à Estrada da Boiúna, na Taquara.

Fonte: Elaboração própria

4 Considerações Finais

Os resultados obtidos confirmam o desempenho promissor da arquitetura *U-Net* para a identificação de favelas em imagens de satélite. De modo geral, o modelo treinado com *data augmentation* apresentou desempenho superior nas métricas de *IoU*, *F1-Score* e Precisão, indicando maior qualidade na sobreposição das áreas segmentadas, melhor equilíbrio entre erros de classificação e maior controle sobre a ocorrência de falsos positivos. Em contrapartida, esse modelo registrou valores ligeiramente inferiores de Revocação em comparação ao modelo treinado sem aumento de dados, evidenciando um comportamento mais conservador na identificação das áreas de interesse.

Esse resultado reflete o *trade-off* clássico entre Precisão e Revocação, no qual o *data augmentation* contribui para tornar o modelo mais seletivo, reduzindo classificações incorretas (falsos positivos), ainda que à custa da perda de algumas detecções verdadeiras. Tal característica pode ser interpretada como desejável em aplicações nas quais a redução de falsos positivos é prioritária, como em análises urbanas e planejamento territorial. Ademais, as análises quantitativas e qualitativas indicam que o uso de *data augmentation* favorece uma segmentação mais estável e espacialmente coerente, especialmente em áreas com maior heterogeneidade

morfológica, reforçando sua relevância como estratégia para aumentar a capacidade de generalização do modelo.

Apesar dos avanços observados, desafios relevantes ainda persistem. O modelo treinado com *data augmentation*, embora apresente maior Precisão, demonstra uma tendência mais conservadora na identificação de áreas de favela, reduzindo a ocorrência de falsos positivos, mas, em alguns casos, deixando de detectar áreas legítimas, especialmente aquelas de pequena extensão ou inseridas em contextos visuais complexos. Por outro lado, o modelo treinado sem aumento de dados, ao apresentar maior Revocação, mostrou-se mais sensível à detecção dessas áreas, porém com maior propensão a classificar erroneamente regiões que não correspondem a esse tipo de ocupação.

Em ambos os casos, erros recorrentes estiveram associados a áreas que apresentam características visuais semelhantes às de favelas, como terraços residenciais, condomínios de pequeno porte e zonas urbanas degradadas. Ademais, a identificação de favelas de pequena extensão, fragmentadas pela grade de recorte ou inseridas em ambientes de baixo contraste visual, mostrou-se particularmente desafiadora, evidenciando limitações na distinção dessas áreas em função da ambiguidade de suas características espectrais e espaciais.

Para superar essas limitações e aprimorar o modelo, sugere-se as seguintes direções para trabalhos futuros:

- 1 Incorporação de bandas espectrais adicionais, como Infravermelho Próximo (NIR) e Infravermelho de Ondas Curtas (SWIR), que podem capturar informações sobre vigor vegetativo, umidade e materiais construtivos, auxiliando na distinção entre favelas e outras tipologias urbanas.
- 2 Refinamento do esquema de classes, com a subdivisão da ampla categoria "não favela" em classes semanticamente distintas. A criação de categorias específicas para tipologias urbanas recorrentemente confundidas — como conjuntos residenciais formais, edificações comerciais, galpões industriais e coberturas de terraço — permitiria ao modelo aprender representações mais discriminativas. Essa maior granularidade na definição do objetivo de aprendizado tem potencial para reduzir significativamente os falsos positivos, proporcionando uma compreensão mais estratificada e contextual do tecido urbano.
- 3 Avaliação de arquiteturas de rede neural mais avançadas, como HRNet, transformers para visão computacional (e.g., SegFormer, Mask2Former) e métodos de Aprendizado Autossupervisionado (SSL), que podem melhorar a capacidade de representação de características espaciais e texturais, especialmente em cenários com dados anotados limitados.
- 4 Exploração de estratégias de pós-processamento para refinamento de máscaras, como técnicas de suavização de bordas, fusão de segmentações e métodos baseados em geometria, visando corrigir incoerências espaciais e reduzir ruídos nas predições.
- 5 Integração de dados contextuais e socioeconômicos em nível de setor censitário, que podem oferecer insights adicionais para a diferenciação de áreas urbanas de baixa renda com morfologias semelhantes.
- 6 Revisão colaborativa do processo de anotação, envolvendo múltiplos especialistas na geração das máscaras de referência. Como a interpretação de características visuais que definem uma favela nem sempre é consensual, a consolidação das anotações por meio de operações de união ou interseção pode produzir máscaras de referência mais consistentes e robustas. Essa abordagem permite capturar tanto uma definição mais abrangente (união) quanto uma mais conservadora (interseção) do fenômeno, reduzindo vieses subjetivos individuais e criando bases de treinamento que melhoram a capacidade de generalização do modelo para diferentes contextos interpretativos.

7 Ampliação e diversificação das técnicas de *data augmentation*, utilizando bibliotecas como Albumentations para introduzir transformações mais complexas e estocásticas, aumentando a variabilidade do conjunto de treinamento e a robustez do modelo a diferentes condições de imagem.

A implementação dessas propostas tem potencial para elevar significativamente a acurácia, a confiabilidade e a aplicabilidade de modelos de segmentação semântica no mapeamento de assentamentos precários, contribuindo para políticas urbanas mais assertivas e baseadas em evidências.

Referências

ABASCAL, A. *et al.* Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. **Journal of Environmental Management**, v. 95, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0198971522000643>. Acesso em: 12 dez. 2023.

ALRASHEEDI, K.; DEWAN, A.; EL-MOWAFY, A. Using local knowledge and remote sensing in the identification of informal settlements in riyadh city, Saudi Arabia. **Remote Sensing**, v. 15, n. 15, p. 3895, 2021. Disponível em: <https://www.mdpi.com/2072-4292/15/15/3895>. Acesso em: 25 set. 2023.

BUSLAEV, A. *et al.* Albumentations: fast and flexible image augmentations. **arXiv preprint arXiv: 1809.06839**. 2018. Disponível em: <https://arxiv.org/pdf/1809.06839>. Acesso em: 16 jan. 2024.

CINNAMON, J.; NOTH, T. Spatiotemporal development of informal settlements in cape town, 2000 to 2020: An open data approach. **Habitat International**, v. 112, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0197397523000139>. Acesso em: 11 abr. 2024.

DATA.RIO. Limite Favelas 2019. 2019. Disponível em: <https://www.data.rio/datasets/limite-favelas-2019/explore>. Acesso em: 20 dez. 2023.

DATA.RIO. Limite Áreas de Planejamento (AP). 2023. Disponível em: https://www.data.rio/datasets/b9e30861acfe4bea947e6278a6b30ce3_1/explore. Acesso em: 20 dez. 2023.

DONG, Shiwei. Spatial Stratification Method for the Sampling Design of LULC Classification Accuracy Assessment: A Case Study in Beijing, China. **Remote Sensing**, v. 14, n. 4, 2022. Disponível em: <https://www.mdpi.com/2072-4292/14/4/865>. Acesso em: 03 fev. 2024.

GHAFFARIAN, S.; EMTEHANI, S. Monitoring urban deprived areas with remote sensing and machine learning in case of disaster recovery. **Urban Science**, v. 9, n. 4, 2021. Disponível em: <https://www.mdpi.com/2225-1154/9/4/58>. Acesso em: 12 dez. 2023.

IBGE. Aglomerados Subnormais 2019: Classificação preliminar e informações de saúde para o enfrentamento à COVID-19. 2020. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101717_notas_tecnicas.pdf. Acesso em: 06 dez. 2023.

IBGE. Favelas e Comunidades Urbanas: IBGE muda denominação dos aglomerados subnormais. 2024. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/38962-favelas-e-comunidades-urbanas-ibge-muda-denominacao-dos-aglomerados-subnormais>. Acesso em: 05 nov. 2025.

KEMPER, T. *et al.* Towards an automated monitoring of human settlements in south africa using high resolution spot satellite imagery. **International Society for Photogrammetry and Remote Sensing**. v. XL-7/W3, [S.n.], 2015. Disponível em:

<https://core.ac.uk/download/pdf/38631179.pdf>. Acesso em: 4 mar. 2024

LAKSHMANAN, Valliappa; GÖRNER, Martin; GILLARD, Ryan. **Practical machine learning for computer vision**. " O'Reilly Media, Inc.", 2021. O'Reilly Media.2021

LING, Antony. Cortiços eram melhores que as favelas. **Caos Planejado**, 2018. Disponível em: <https://caosplanejado.com/corticoseeram-melhores-que-favelas/>. Acesso em: 23 dez. 2023.

LU, W. *et al.* A Geoscience-Aware Network (GASlumNet) Combining UNet and ConvNeXt for Slum Mapping. **Remote Sensing**, v. 16, n. 2, p. 260, 2021. Disponível em: <https://www.mdpi.com/2072-4292/16/2/260>. Acesso em: 14 mai. 2024.

MAIYA, S. R.; BABU, S. C. Slum segmentation and change detection: A deep learning approach. **arXiv preprint arXiv**: 1811.07896. 2018. Disponível em: <https://arxiv.org/pdf/1811.07896>. Acesso em: 13 mar. 2024.

MARINS, Paulo. **História da Vida Privada no Brasil**: Habitação e Vizinhança: Limites da Privacidade no Surgimento das Metrópoles Brasileiras. 3. ed. São Paulo: Companhia Das Letras, 1998. P. 86.

OLIVEIRA, L. T. *et al.* Capturing deprived areas using unsupervised machine learning and open data: a case study in São Paulo, Brazil. **European Journal of Remote Sensing**, 2023. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/22797254.2023.2214690>. Acesso em: 5 mar. 2024.

ONU BRASIL. **ONU-Habitat**: população mundial será 68% urbana até 2050. NAÇÕES UNIDAS BRASIL. 2022. Disponível em: <https://brasil.un.org/pt-br/188520-onu-habitat-popula%C3%A7%C3%A3o-mundial-ser%C3%A1-68-urbana-at%C3%A9-2050>. Acesso em: 17 set. 2023.

ONU BRASIL. **Os objetivos de desenvolvimento sustentável no Brasil**. NAÇÕES UNIDAS BRASIL. 2023. Disponível em: <https://brasil.un.org/pt-br/sdgs/11>. Acesso em: 17 set. 2023.

PREFEITURA DA CIDADE DO RIO DE JANEIRO. **Conselho Estratégico de Informações da Cidade**: Atas de Reuniões. 2012. Disponível em: <https://www.rio.rj.gov.br/documents/91329/1f8a19d9-91d6-430d-81f4-52081055114e>. Acesso em: 13 nov. 2025.

PREFEITURA DA CIDADE DO RIO DE JANEIRO. **Painel Rio**: Conheça mais o rio de hoje para construir o rio de amanhã. 2025. Disponível em: <https://pds-pcrj.hub.arcgis.com/pages/unidades>. Acesso em: 12 nov. 2025.

PROVOST, Foster; FAWCETT, Tom. **Data Science para Negócios**: O que você precisa saber sobre mineração de dados e pensamento analítico de dados. Rio de Janeiro: Alta Books, 2016.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. **arXiv preprint arXiv**: 1505.04597, 2015. Disponível em: <https://arxiv.org/pdf/1505.04597>. Acesso em: 16 jan. 2024.

SILVA, Pedro; BIANCHINI, Zélia; DIAS, Antonio. **Amostragem: Teoria e Prática Usando R**. Rio de Janeiro: [S.n.], 2023. Disponível em: <https://amostragemcomr.github.io/livro/index.html>. Acesso em: 2 feb. 2026.

USGS (United States Geological Survey). **The Universal Transverse Mercator (UTM) Grid**. [S.l.: s.n.], 2001. P. 2. Documento online. Disponível em: <https://pubs.usgs.gov/fs/2001/0077/report.pdf>. Acesso em: 10 jan. 2024.

WURM, M. *et al.* Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 150, 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0924271619300383>. Acesso em: 11 dez. 2023

Sobre os Autores

Jedielso Sales de Souza é graduado em Estatística pela Escola Nacional de Ciências Estatísticas (ENCE). Atua no Instituto Brasileiro de Geografia e Estatística (IBGE) como Chefe da Agência Centro da Superintendência Estadual do Rio de Janeiro, sendo responsável pelo gerenciamento e treinamento de equipes envolvidas na coleta de diversas pesquisas estatísticas. No Censo Demográfico de 2022, exerceu a função de coordenador de área na região central da cidade do Rio de Janeiro, onde adquiriu experiência no mapeamento de favelas e comunidades urbanas. Sua atuação integra interesses em estatística, análise de dados espaciais e estudos sobre favelas.

Andrea Diniz da Silva é professora e pesquisadora nos programas de graduação e de pós-graduação da Escola Nacional de Ciências Estatísticas - ENCE/IBGE; líder do grupo de pesquisas "Big Data e Estatísticas Públicas"; pesquisadora dos grupos de pesquisas "DHPJS - Direitos Humanos, Poder Judiciário e Sociedade" e "CAST - Computational Agriculture Statistics Laboratory". Bacharel em estatística pela Universidade do Estado do Rio de Janeiro - UERJ, mestre em Estudos Populacionais e Pesquisas Sociais e doutora em População, Território e Estatísticas Públicas, ambos pela ENCE. Fez doutorado sanduíche na University of Wollongong, na Austrália. Trabalha com métodos para pesquisas e uso de big data para a produção de estatísticas públicas. Ocupa o cargo de chefe de relações internacionais e de coordenadora de GT sobre política de inovação, no IBGE.

Ian Monteiro Nunes é Engenheiro de Computação com mestrado e doutorado em Ciência de Dados pelo Departamento de Informática da PUC-Rio. Atualmente, é o Head da Gerência de Inteligência em Dados e Inovação do IBGE, área pioneira voltada para P&D e modernização do Censo Agropecuário, e professor de Ciência de Dados na Escola Nacional de Ciências Estatísticas (ENCE). Sua atuação inclui a coordenação do projeto de R&D "Representation Learning for Crop Segmentation in Brazil" no LNCC e a participação no grupo de pesquisa Big Data e Estatísticas Oficiais. Com sólida experiência na fronteira entre a academia e a gestão pública, Ian lidera iniciativas de transformação digital utilizando Visão Computacional, Sensoriamento Remoto e Aprendizado de Máquina. Suas áreas de pesquisa abrangem segmentação de imagens, deep learning, adaptação de domínio e open-set recognition.

Contribuições dos Autores

Conceituação, [J.S.S]; metodologia, [J.S.S., A.D.S., I.M.N]; software [J.S.S]; validação, [J.S.S]; análise formal, [J.S.S]; investigação, [J.S.S]; recursos, [J.S.S]; curadoria de dados, [J.S.S]; redação—preparação do rascunho original, [J.S.S]; redação—revisão e edição [J.S.S., A.D.S., I.M.N]; visualização, [J.S.S]; supervisão, [A.D.S., I.M.N]. Todos os autores leram e concordaram com a versão publicada do manuscrito.

Disponibilidade de Dados

Os dados para esta pesquisa estão disponíveis em: <https://drive.google.com/drive/folders/10j4KrWzsuuQuK8yODZRRhCscWvNig17J?usp=sharing>

Conflitos de Interesse

Os autores declaram não haver conflitos de interesse.

Sobre a Coleção Estudos Cariocas

A Coleção Estudos Cariocas (ISSN 1984-7203) é uma publicação de estudos e pesquisas sobre o Município do Rio de Janeiro, vinculada ao Instituto Pereira Passos (IPP) da Secretaria Municipal da Casa Civil da Prefeitura do Rio de Janeiro.

Seu objetivo é divulgar a produção técnico-científica sobre temas relacionados à cidade do Rio de Janeiro, bem como sua vinculação metropolitana e em contextos regionais, nacionais e internacionais. Está aberta a quaisquer pesquisadores (sejam eles servidores municipais ou não), abrangendo áreas diversas - sempre que atendam, parcial ou integralmente, o recorte espacial da cidade do Rio de Janeiro.

Os artigos também necessitam guardar coerência com os objetivos do Instituto, a saber:

1. Promover e coordenar a intervenção pública sobre o espaço urbano do Município;
2. Prover e integrar as atividades do sistema de informações geográficas, cartográficas, monográficas e dados estatísticos da Cidade;
3. Subsidiar a fixação das diretrizes básicas ao desenvolvimento socioeconômico do Município.

Especial ênfase será dada no tocante à articulação dos artigos à proposta de desenvolvimento econômico da cidade. Desse modo, espera-se que os artigos multidisciplinares submetidos à revista respondam às necessidades de desenvolvimento urbano do Rio de Janeiro.