

# Modelo de Predição de Perda de Propagação do Sinal de Ondas Sub-6GHz Utilizando Aprendizado Profundo e Dados Visuais Geoespaciais

Caio P. Galdino<sup>1</sup>, Pedro Henrique Cruz Caminha<sup>1</sup>, Rodrigo de Souza Couto<sup>1</sup>

<sup>1</sup>GTA/DEL – POLI/PEE – COPPE  
Universidade Federal do Rio de Janeiro (UFRJ)

{galdino, cruz, rodrigo}@gta.ufrj.br

**Abstract.** *Efficient planning for radiocommunications, particularly at sub-6GHz frequencies, requires accurate path loss estimation. This work proposes a Deep Learning architecture that integrates numerical and 2.5D geospatial visual data, applying a Partial Self-Attention (PSA) mechanism to weigh the impact of obstacles along the signal path. The model is trained and validated using real-world measurement data from German cities, achieving a Root Mean Square Error (RMSE) of 1.97 dB and a Mean Absolute Error (MAE) of 1.18 dB, outperforming traditional approaches. Additionally, transfer learning from the city of Moers to Frankfurt demonstrates the method's robustness, yielding an RMSE of 2.41 dB and an MAE of 1.89 dB.*

**Resumo.** *O planejamento eficiente de radiocomunicações, especialmente em frequências sub-6GHz, exige estimativas precisas de perda de propagação. Este trabalho propõe uma arquitetura de Aprendizado Profundo que integra dados numéricos e visuais geoespaciais 2.5D, aplicando o mecanismo de Atenção Parcial para ponderar a relação de obstáculos ao longo do trajeto do sinal. O modelo é treinado e validado com dados reais de medição em cidades alemãs, atingindo um erro quadrático médio (RMSE) de 1,97 dB e Erro Médio Absoluto (MAE) de 1,18 dB, que supera as abordagens tradicionais. Adicionalmente, a transferência de aprendizado na cidade de Moers para Frankfurt demonstra a robustez do método, com RMSE de 2,41 e MAE de 1,89 dB.*

## 1. Introdução

Com a evolução das redes celulares para o 5G e a iminência do 6G, observa-se uma utilização crescente de faixas de frequência mais elevadas. Esse cenário impõe novos desafios para o planejamento de cobertura, dado que sinais nessas frequências sofrem atenuação severa por obstruções físicas, comportando-se de maneira quase óptica. Serviços críticos do 5G, como eMBB (*Enhanced Mobile Broadband*), URLLC (*Ultra-Reliable Low Latency Communication*) e mMTC (*Massive Machine Type Communication*), dependem estritamente de uma Qualidade de Serviço (QoS) garantida. Nesse contexto, o gerenciamento eficiente do espectro eletromagnético torna-se necessário. Métricas como o RSRP (*Reference Signal Received Power*) e o *Path Loss* (perda de percurso) são fundamentais para quantificar como fatores ambientais (relevo, edificações e vegetação) degradam o sinal recebido pelo usuário.

Historicamente, a estimativa de cobertura baseia-se em modelos de propagação que podem ser categorizados, majoritariamente, em duas classes: modelos empíricos e modelos determinísticos [Vasudevan and Yuksel 2024]. Os modelos empíricos (ou estocásticos), como Okumura-Hata e COST-231, baseiam-se em equações estatísticas derivadas de campanhas de medição exaustivas. Embora sejam computacionalmente eficientes, eles falham em capturar as especificidades da geometria local, resultando em baixa precisão em ambientes urbanos complexos (*Urban Micro-cells*). Por outro lado, modelos determinísticos, baseados em *Ray-Tracing* (traçado de raios), simulam fisicamente a interação da onda com o ambiente 3D. Apesar de oferecerem alta fidelidade, seu custo computacional é proibitivo para o planejamento de redes em larga escala, exigindo bases de dados vetoriais 3D detalhadas e alto tempo de processamento [Lee et al. 2019]. Recentemente, a aplicação de técnicas de Aprendizado Profundo (*Deep Learning*) emergiu como uma alternativa promissora para superar o compromisso entre precisão e custo computacional. Redes Neurais Convolucionais (CNNs), tradicionalmente utilizadas em visão computacional, demonstram capacidade de extrair características complexas de ambientes de propagação a partir de dados visuais, como imagens de satélite e mapas de elevação [Waheed et al. 2022a, Zeng et al. 2024]. Diferentemente dos métodos empíricos, modelos baseados em IA conseguem “aprender” a influência de obstáculos específicos no percurso do sinal, aproximando-se da precisão do *Ray-Tracing* com um tempo de inferência significativamente menor.

Arquiteturas baseadas puramente em CNNs muitas vezes tratam o ambiente de forma homogênea, sem distinguir adequadamente quais obstáculos interagem diretamente com a linha de visada do sinal. Este trabalho propõe uma abordagem híbrida e multimodal que integra a física dos modelos clássicos com a capacidade de representação de dados visuais geoespaciais 2.5D (elevação e semântica). A principal contribuição reside na aplicação do mecanismo de Atenção Parcial (*Partial Self-Attention* - PSA) para ponderar dinamicamente a relevância dos obstáculos ao longo do trajeto, permitindo uma predição de perda de propagação mais precisa, validada com dados reais de redes operacionais. A metodologia proposta neste trabalho baseia-se no estado da arte de modelos simultaneamente **multimodais** e **híbridos**. A abordagem é classificada como multimodal por processar dados de naturezas distintas, como imagens de satélite e parâmetros numéricos de rede. O caráter híbrido deve-se à utilização do Aprendizado Profundo para refinar modelos empíricos de base, como o Okumura-Hata, em vez de substituir integralmente a modelagem física.

## 2. Trabalhos Relacionados

Na literatura, fundamenta-se que as abordagens clássicas impõem uma dicotomia a respeito da predição da perda de propagação. Vasudevan *et al.* [Vasudevan and Yuksel 2024] resumizam o problema da perda de propagação do sinal, elencando e diferenciando os principais e recentes trabalhos relacionados ao tema, introduzindo métodos de tratamentos de dados, assim como as principais equações de perda de percurso utilizadas pelos autores, dividindo a literatura entre modelos empíricos, que são computacionalmente eficientes, mas ignoram a geometria específica do local, e modelos determinísticos (como *Ray Tracing*), que oferecem alta precisão ao custo de uma carga computacional proibitiva para o tempo de resposta exigido para operações em tempo real.

O estudo de Cavalcanti *et al.* [Cavalcanti et al. 2025] demonstra que o treinamento

de modelos unimodais, restritos a uma única natureza de dado, sejam apenas parâmetros numéricos ou apenas imagens, apresenta limitações intrínsecas de desempenho. Essas lacunas são mitigadas pelos modelos multimodais, que operam através da fusão de diferentes fontes de informação (visual e tabular) para enriquecer a representação do ambiente. Distintamente, os métodos híbridos não se caracterizam pela mistura de dados, mas sim pela integração de paradigmas de modelagem: eles combinam modelos teóricos (empíricos ou determinísticos) com algoritmos de aprendizado de máquina. Nessa abordagem, a IA é projetada para prever o erro residual (a diferença entre a física teórica e a realidade medida), resultando em uma acurácia superior em cenários de alta complexidade, como o ambiente urbano.

Waheed *et al.* [Waheed et al. 2022a] propõem uma arquitetura multimodal e híbrida projetada para superar o custo computacional do *Ray Tracing* e a dependência de mapas 3D. O trabalho reforça a estratégia de combinar dados numéricos com imagens de satélite do entorno do receptor para a extração automática de características ambientais. A pesquisa valida a técnica de transferência de aprendizado (*Transfer Learning*): o modelo generaliza o conhecimento adquirido em uma região de rede LTE para outra com características distintas, o que reduz o tempo de treinamento para novos cenários sem prejuízo à precisão da predição. Em desdobramentos dessa pesquisa, Waheed *et al.* [Waheed et al. 2022b] e Zeng *et al.* [Zeng et al. 2024] incorporam a segmentação semântica ao processamento das imagens. A identificação explícita de características ambientais mostra-se superior ao uso de imagens brutas, pois permite ao modelo distinguir classes de obstáculos no terreno. A riqueza semântica da entrada possibilita reduzir a complexidade e o tamanho da rede neural, superando a acurácia dos trabalhos anteriores.

A importância da geometria 2.5D/3D e do recorte específico da imagem do caminho entre transmissor e receptor é enfatizada por autores como Lee *et al.* [Lee et al. 2019], Sani *et al.* [Sani et al. 2022] e Yeas *et al.* [Yeaser and Hassan 2025]. Lee *et al.* [Lee et al. 2019] validam que a consideração concreta de características topográficas, por meio da codificação da altura das edificações e da densidade do ambiente, beneficia a modelagem em frequências de ondas milimétricas. Essa representação permite inferir o Expoente de Perda de Percurso com precisão comparável ao *Ray Tracing*, mas com tempo de inferência reduzido, viabilizando aplicações de resposta rápida.

A pesquisa conduzida por Sani *et al.* [Sani et al. 2022] compara o desempenho de descritores visuais manuais (*hand-crafted*) contra aqueles gerados por CNNs, e conclui que a abordagem manual resulta em degradação da acurácia. A análise de explicabilidade via shap (*SHapley Additive exPlanations*) revela que as características ambientais extraídas pela CNN, a partir de imagens de satélite dispostas ao longo do trajeto, constituem os fatores de maior contribuição para a decisão do modelo, superando a relevância isolada das variáveis numéricas tradicionais.

O trabalho de Yeas *et al.* [Yeaser and Hassan 2025] consolida a validação de arquiteturas de Aprendizado Profundo frente aos métodos clássicos. O estudo explora um conjunto diversificado de redes neurais, integrando CNN para padrões espaciais e modelos recorrentes LSTM (*Long Short-Term Memory*) e GRU (*Gated Recurrent Unit*) para dependências sequenciais. A capacidade dessas arquiteturas de modelar interações não-lineares complexas mostra-se essencial para cenários nas faixas de frequência sub-6GHz, nos quais os métodos tradicionais falham em refletir mudanças nas condições ambientais.

No contexto de detecção de objetos, o artigo de Wang *et al.* [Wang et al. 2024] propõe o *Partial Self-Attention*. Essa técnica foi introduzida recentemente, especificamente na arquitetura YOLOv10, para solucionar o problema da complexidade quadrática dos mecanismos de atenção tradicionais, resolvendo o dilema entre custo computacional e capacidade de representação global, tornando-se ideal para a aplicação de predição rápida.

As abordagens atuais da literatura, contudo, ainda não exploram mecanismos de atenção para ponderar elementos ao longo do trajeto e aprender relações complexas dos objetos que compõem a trajetória do sinal. A proposta deste trabalho preenche essa lacuna por meio de uma arquitetura que unifica a geometria do problema e o foco no percurso com a informação semântica. A inovação reside na adição de mecanismos de atenção sequencial para emular a física da propagação através dos obstáculos identificados.

### 3. Conjunto de Dados e Pré-Processamento

A estratégia de treinamento e validação deste trabalho fundamenta-se no uso de dados empíricos reais, em contraste com abordagens baseadas puramente em simulações sintéticas. Utiliza-se o **ITU-R SG3 Databank** (Banco de Dados da Comissão de Estudos 3 da União Internacional de Telecomunicações)<sup>1</sup>, reconhecido pela diversidade e rigor de suas campanhas de medição. Para este estudo, são selecionadas campanhas de medição da perda de propagação do sinal do **ITU-R SG3 Databank** realizadas em cenários urbanos densos na Alemanha nas cidades de Moers, Friedrichshafen, Frankfurt e Mannheim. Esta escolha justifica-se pela necessidade de capturar a complexidade da propagação em ambientes reais, sujeitos a efeitos de cânions urbanos (*street canyons*), difração em arestas de edificações e espalhamento, fenômenos que modelos idealizados frequentemente subestimam, além disso, a Alemanha foi escolhida pela disponibilidade de um banco de dados aberto, maduro e com geolocalização absoluta precisa em ambas as pontas do enlace (Tx e Rx) que são necessárias para calcular as imagens de entrada do modelo. As campanhas de Friedrichshafen são feitas em frequências de 2300 MHz, com o intuito de treinar um modelo em faixas de frequências o mais próximo dos trabalhos relacionados para fins de comparações verossímeis de ambiente e parâmetros de operação. Trabalhos como o de Waheed *et al.* [Waheed et al. 2022a, Waheed et al. 2022b] utilizam dados de 2.100 MHz nas cidades de Cairo e Giza, assim como Zeng *et al.* [Zeng et al. 2024] propõem seu estudo nas faixas de 811 MHz, 2.650 MHz em Wangjiang. A cidade de Moers e demais escolhidas operam, respectivamente, nas frequências de 5,45 GHz e 5,85 GHz, nas quais a interação da onda com a geometria dos obstáculos torna-se crítica, com a onda apresentando um comportamento ondulatório mais complexo do que em frequências mais baixas. A Tabela 1 mostra o conjunto de dados das cidades e suas respectivas frequências dos dados coletados.

Um critério de inclusão para a seleção dos dados foi a disponibilidade de geolocalização precisa. Ao contrário de conjuntos de dados históricos que fornecem apenas a distância escalar ( $d$ ) entre transmissor ( $Tx$ ) e receptor ( $Rx$ ), a arquitetura multimodal proposta exige as coordenadas geográficas absolutas (latitude e longitude) de ambas as pontas do enlace. Essa informação é pré-requisito para a aquisição das imagens de satélite correspondentes ao trajeto do sinal, viabilizando a extração de características visuais e a aplicação do filtro de remoção e suavização de dados ruidosos.

---

<sup>1</sup><https://www.itu.int/en/ITU-R/study-groups/rsg3/Pages/dtbank-dbsg3.aspx>

Tabela 1. Resumo do Conjunto de Dados e Amostras por Cidade.

Cidade	Frequência (MHz)	Amostras
Friedrichshafen	2.300	48.283
Mannheim	5.850	88.294
Moers	5.450 e 5.850	75.708
Frankfurt	5.850	4.235

### 3.1. Tratamento de Dados Numéricos

Dados brutos de medição são inerentemente ruidosos, sujeitos a erros de instrumentação, falhas de GPS e efeitos de desvanecimento rápido (*fast fading*). Para garantir a integridade do treinamento, aplica-se um protocolo de sanitização e suavização inspirado na metodologia de Zeng *et al.* [Zeng et al. 2024]. O processo consiste em quatro etapas:

1. **Remoção de Dados Incompletos:** Descarte de amostras contendo valores nulos ou corrompidos em campos críticos (coordenadas, frequência ou potência).
2. **Filtragem de Faixa Dinâmica:** Remoção de registros de perda fora da sensibilidade típica de receptores comerciais (abaixo de -140 dBm ou acima de -40 dBm), considerados ruído ou saturação.
3. **Consistência Geométrica:** Exclusão de amostras com distâncias  $Tx - Rx$  irrealistas para microcélulas urbanas ( $< 10$  m ou  $> 5$  km).
4. **Suavização Espacial (*Spatial Averaging*):** Filtragem espacial baseada em coordenadas absolutas. Devido à ausência de marcadores temporais (*timestamps*) nas medições brutas, não é possível aplicar médias temporais tradicionais. Adotou-se, portanto, uma abordagem análoga ao algoritmo de grade proposto por Zeng *et al.* [Zeng et al. 2024]. As medições são agrupadas por proximidade geográfica, calculando-se a média da perda para pontos dentro de um raio de coerência espacial. Este processo mitiga variações estocásticas do desvanecimento rápido, isolando o componente de desvanecimento lento (*shadowing*), o alvo de predição do modelo.

### 3.2. Aquisição e Processamento de Dados Visuais

A componente visual da arquitetura multimodal exige um tratamento específico para garantir a correspondência entre a resolução da imagem e a capacidade de entrada das redes neurais. As imagens de satélite são obtidas através da **Google Maps Static API**<sup>2</sup>, centradas no ponto médio entre o transmissor e o receptor para cada enlace validado.

O pipeline de processamento visual adotado neste trabalho segue uma estratégia de redução progressiva de dimensionalidade (*downsampling*) para preservar informações contextuais sem comprometer o custo computacional:

1. **Aquisição (1024×1024):** As imagens brutas são baixadas em alta resolução (1024 × 1024 pixels) e fator de zoom calibrado para cobrir a área de interesse do enlace, garantindo que detalhes finos da malha urbana sejam capturados.
2. **Segmentação Semântica (512×512):** Para a etapa de segmentação (descrita na Subseção 4.2), as imagens de saída da segmentação são redimensionadas para

<sup>2</sup><https://developers.google.com/maps/documentation/maps-static>

512 × 512 pixels. A resolução representa um compromisso ideal para a arquitetura U-Net, mantendo a fidelidade das bordas de edifícios e vegetação essencial para a inferência correta da classe da área segmentada, enquanto otimiza o uso de memória da GPU para as imagens de saída geradas.

3. **Entrada do Modelo Híbrido (224×224):** Os mapas de elevação e as máscaras de segmentação resultantes foram redimensionados de 512 × 512 para 224 × 224 pixels. Esta dimensão é o padrão de entrada da rede *backbone* AlexNet utilizado na extração de características, garantindo a compatibilidade com os pesos pré-treinados e reduzindo a complexidade da inferência final.

#### 4. Arquitetura do Modelo

A arquitetura, ilustrada na Figura 1, consiste em um modelo híbrido que funde dados visuais do ambiente com parâmetros numéricos do sistema, baseado no trabalho de Waheed *et al.* [Waheed et al. 2022a]. Ela é dividida em dois ramos principais: um para tratamento de dados numéricos (representado no ramo superior da Figura 1) e um para o tratamento de dados visuais ((representado no ramo inferior da Figura 1)). Cada ramo é detalhado nas seguintes seções.

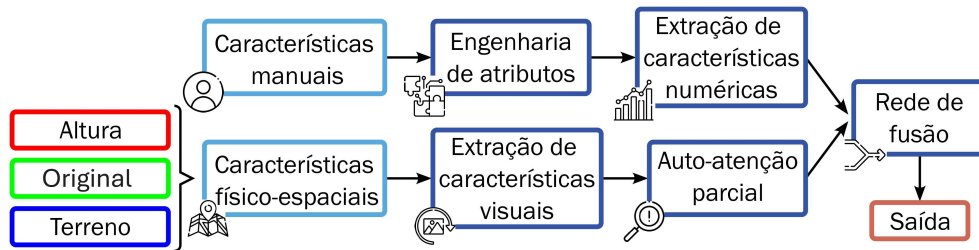


Figura 1. Diagrama do Modelo de Perda de Propagação.

##### 4.1. Processamento de Dados Numéricos

O ramo superior da arquitetura proposta na Figura 1 dedica-se ao processamento de dados estruturados e à incorporação de conhecimento de domínio, estabelecendo o caráter híbrido do sistema. Seguindo a metodologia de características manuais (*handcrafted features*) validada por Waheed *et al.* [Waheed et al. 2022a], o modelo é alimentado com um conjunto de variáveis determinísticas essenciais para a física da propagação. O vetor de entrada é composto por coordenadas geográficas, parâmetros geométricos do enlace, distância euclidiana, ângulo de azimute entre transmissor e receptor, e estimativa de perda de percurso calculada pelo modelo empírico Hata-Okumura conforme a equação a seguir:

$$L_p(\text{dB}) = 69.55 + 26.16 \log_{10}(f) - 13.82 \log_{10}(h_b) - a(h_m) + [44.9 - 6.55 \log_{10}(h_b)] \log_{10}(d). \quad (1)$$

O termo  $a(h_m)$  representa o fator de correção para a altura da antena móvel e varia conforme o tipo de ambiente urbano.

Para cidades de pequeno e médio porte, o fator de correção é dado por:

$$a(h_m) = (1.1 \log_{10}(f) - 0.7)h_m - (1.56 \log_{10}(f) - 0.8). \quad (2)$$

Para grandes cidades (considerando frequências  $f \geq 400$  MHz), utiliza-se:

$$a(h_m) = 3.2(\log_{10}(11.75h_m))^2 - 4.97, \quad (3)$$

tal que:

- $L_p$ : Perda de propagação (*Path Loss*) em decibels (dB);
- $f$ : Frequência de operação em MHz;
- $h_b$ : Altura da antena da estação base (transmissor) em metros;
- $h_m$ : Altura da antena móvel (receptor) em metros;
- $d$ : Distância entre transmissor e receptor em quilômetros (km);
- $a(h_m)$ : Fator de correção dependente da altura da antena móvel e da densidade urbana.

A inclusão desta estimativa empírica como uma *feature* auxiliar é estratégica, pois direciona a rede neural para o aprendizado do erro residual do modelo clássico em vez da função de propagação integral, acelerando a convergência e garantindo que as previsões mantenham coerência com os princípios físicos do problema [Waheed et al. 2022a]. Antes da etapa de previsão pelo modelo, os dados brutos são normalizados para mitigar a disparidade de escalas entre grandezas heterogêneas (como frequência em MHz e atenuação em dB), evitando instabilidades no gradiente durante o treinamento. Subsequentemente, os dados normalizados alimentam o componente representado na Figura 1 por “Extração de Características Numéricas”, constituído por uma rede MLP (*Multi Layer Perceptron*). Conforme a arquitetura de referência [Waheed et al. 2022a], este bloco é composto por uma sequência de camadas densas com ativação não-linear (ReLU). Sua função primária não é realizar a regressão final, mas atuar como um codificador (*encoder*), projetando as características escalares de entrada em um vetor de características latentes (*embedding*) de alta dimensionalidade. Este processo transforma o contexto numérico e físico do sistema em uma representação vetorial compatível para a futura fusão com as características visuais extraídas pelo segundo ramo da arquitetura.

## 4.2. Representação do Ambiente e Codificação Visual

Paralelamente ao processamento numérico, o segundo ramo da arquitetura dedica-se à extração de padrões espaciais representada na Figura 1 por “Extração de Características Visuais”. Diferentemente de abordagens que utilizam imagens de satélite brutas ou apenas a imagem de saída da segmentação semântica, este trabalho adota a construção de um tensor sintético de três canais (*2.5D*). A estratégia de codificação inspira-se na metodologia estabelecida por Lee *et al.* [Lee et al. 2019], que demonstra a eficácia de embutir informações geométricas diretamente nos canais RGB para enriquecer a interpretação da rede neural. Para esta proposta, foi utilizada a seguinte configuração dos canais:

- **Canal Vermelho ( $R$ ):** Codifica a elevação do terreno. Os dados são provenientes do modelo digital global ALOS World 3D (AW3D30)<sup>3</sup>. O valor da intensidade do pixel representa a altura da superfície naquele ponto em relação ao nível do

<sup>3</sup><https://portal.opentopography.org/datasetMetadata?otCollectionID=OT.112016.4326.2>

mar, normalizada pelo ponto que possui a maior elevação do cenário capturado. A abordagem utiliza a elevação direta do satélite, conforme a Equação 4:

$$R(x, y) = \frac{H_{ALOS}(x, y)}{H_{ALOS}(x_{max}, y_{max})} \times 255. \quad (4)$$

- **Canal Verde (G):** Armazena a informação visual original da imagem de satélite. Assim, são preservados detalhes da imagem, como a textura de telhados, ruas e vegetação, que auxiliam o modelo a inferir os obstáculos.
- **Canal Azul (B):** Codifica a Segmentação Semântica.

A inovação proposta nesta arquitetura reside na utilização do canal azul para a codificação semântica explícita. Enquanto trabalhos anteriores utilizam este canal para relações espaciais genéricas [Waheed et al. 2022a, Park et al. 2024], ou utilizam apenas a saída da segmentação semântica como entrada do modelo [Waheed et al. 2022b, Zeng et al. 2024], este projeto aplica o resultado de um modelo de segmentação próprio baseado em U-Net e treinado com o conjunto de dados da competição DeepGlobe [Demir et al. 2018] para categorizar o uso do solo. A Tabela 2 detalha o desempenho do modelo de segmentação. A principal métrica considerada para validação foi a acurácia pixel a pixel, visto que o mIoU pode subestimar a performance global ao penalizar classes raras. Como os testes qualitativos demonstraram uma segmentação robusta nos elementos estruturais do cenário urbano (o foco desta aplicação), a acurácia reflete melhor a efetividade do modelo para os objetivos deste estudo. A codificação utiliza faixas de tons específicas no canal azul para representar classes de materiais com propriedades eletromagnéticas distintas, conforme detalhado na Tabela 3. Essa segmentação prévia remove a ambiguidade visual da imagem tradicional, entregando ao extrator de características uma representação da natureza material de cada obstáculo no percurso.

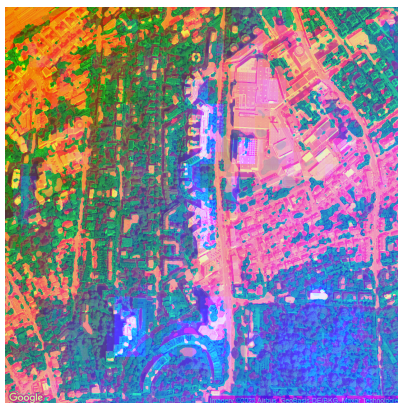
**Tabela 2. Métricas de Desempenho do Modelo de Segmentação (U-Net).**

Métrica	Valor Obtido
Perda	0.9839
Acurácia	82.79%
Mean Intersection over Union (mIoU)	47.42%

**Tabela 3. Mapeamento de Classes Semânticas no Canal Azul.**

Classe Semântica	Intensidade Canal Azul
Urbano	<b>40</b>
Agricultura	<b>80</b>
Pastagem	<b>120</b>
Floresta	<b>160</b>
Água	<b>200</b>
Solo Exposto	<b>240</b>

A Figura 2 ilustra a composição final da imagem resultante 2.5D para a cidade de Frankfurt. Destaca-se como a combinação da elevação ALOS-3D (Vermelho) com a textura visual (Verde) e a semântica (Azul) oferece uma representação rica para o extrator de características visuais que preserva os detalhes da imagem original.



**Figura 2. Configuração Final: ALOS-3D + Canal verde Original + Segmentação Semântica.**

O tensor visual 2.5D multimodal proposto serve, então, como entrada para o *backbone* extrator de características visuais da rede neural convolucional, que transforma a imagem em uma representação latente, preparando o contexto para a aplicação dos mecanismos de atenção na etapa subsequente.

#### **4.3. Backbone Extrator de Características**

O processamento do tensor visual 2.5D é realizado por uma CNN que atua como o *backbone* extrator de características do sistema. Seguindo a metodologia validada por Waheed *et al.* [Waheed et al. 2022a], adota-se a arquitetura AlexNet para esta função devido à sua capacidade comprovada de capturar padrões espaciais complexos em dados de propagação. A escolha garante que os ganhos de performance obtidos nesta pesquisa sejam atribuídos estritamente à modelagem 2.5D e ao mecanismo de Atenção Parcial (PSA) propostos, e não ao poder de extração de um backbone de última geração. Nesta etapa, a rede não processa a imagem como uma fotografia convencional, mas interpreta as relações não-lineares codificadas nos canais de cores definidos. As camadas convolucionais aprendem a correlacionar a intensidade dos canais geométricos (Vermelho e Verde) com as categorias semânticas do canal Azul. Desta forma, o modelo identifica padrões de obstrução complexos, como a distinção entre uma elevação natural do terreno (Canal Vermelho alto + Canal Azul "Solo") e um bloqueio artificial denso (Canal Vermelho alto + Canal Azul "Urbano"). A arquitetura composta por blocos convolucionais, funções de ativação ReLU e camadas de normalização transforma a entrada visual bruta em um conjunto de mapas de características de alto nível. Estes mapas preservam a estrutura espacial da cena, permitindo que a informação sobre onde o obstáculo está localizado no percurso seja mantida para a etapa de atenção subsequente.

#### **4.4. Mecanismo de Atenção: PSA**

A principal inovação arquitetural proposta neste trabalho reside na integração de um módulo de atenção avançado entre o extrator de características e a cabeça de regressão. Embora CNNs sejam eficientes na extração de características locais, elas possuem limitações intrínsecas para capturar dependências de longo alcance e contexto global, fatores cruciais para a modelagem de propagação na qual a obstrução depende da relação espacial entre múltiplos obstáculos distantes entre si que compõem a trajetória

do sinal. Para mitigar esta limitação sem incorrer no custo computacional proibitivo dos mecanismos de *Multi-Head Self-Attention* (MHSA) tradicionais, adota-se o mecanismo de PSA. O funcionamento do PSA baseia-se na divisão estratégica dos canais de características. Em vez de aplicar o cálculo de atenção a todo o tensor (o que gera uma complexidade quadrática), o módulo divide as características em duas partes. A primeira é processada por um bloco de atenção pura, que captura as relações globais (ex: a influência da elevação em um dado ponto em uma sombra em outro ponto da imagem). A segunda parte segue um fluxo simplificado de rede *feed-forward*, preservando informações locais e reduzindo o tempo de processamento em relação à atenção pura. A aplicação deste componente ao problema de regressão de perda de sinal fundamenta-se na hipótese de que a importância de um obstáculo não é estática. Um edifício alto (identificado pelo *backbone*) deve ter um peso maior na predição final se estiver cruzando a linha de visada crítica, e um peso menor se estiver na periferia. O PSA permite que a rede aprenda dinamicamente essa ponderação espacial, focando os recursos computacionais nas características do terreno que efetivamente degradam o sinal, melhorando a generalização do modelo para novos ambientes urbanos, diminuindo a quantidade de dados e tempo de convergência das métricas para a transferência de aprendizado.

#### 4.5. Fusão Multimodal e Regressão Final

A etapa final da arquitetura consiste na integração dos dois fluxos de processamento: o contexto numérico codificado e o contexto visual ponderado pelo mecanismo de atenção para realizar a predição da perda de propagação. Esta integração é executada através de uma estratégia de *Late Fusion* (Fusão Tardia) [Waheed et al. 2022a], na qual os vetores de características latentes resultantes de ambos os ramos são concatenados em um único vetor denso de alta dimensionalidade. Este vetor unificado alimenta uma segunda rede neural denominada Rede de Fusão, o componente regressor final do sistema. Arquiteturalmente, este bloco é constituído por camadas densas (*Fully Connected*) que progressivamente reduzem a dimensionalidade dos dados até um único neurônio de saída com ativação linear, representando o valor estimado da perda de sinal em decibels (*dB*). Embora a arquitetura não imponha uma soma residual rígida, a presença da estimativa do modelo Hata-Okumura no vetor de entrada numérico fornece ao regressor uma “linha de base” física robusta. Consequentemente, a rede neural tende a operar como um mecanismo de refinamento não-linear: em vez de inferir a física de propagação do zero (*tabula rasa*), o modelo aprende a ponderar as características visuais extraídas para ajustar ou corrigir a estimativa empírica base, resultando em uma predição final que combina a estabilidade dos modelos clássicos com o refinamento dos dados geoespaciais da região do Aprendizado Profundo.

### 5. Configuração de Hiperparâmetros e Treinamento

Os experimentos são conduzidos treinando os modelos independentes para os cenários de Friedrichshafen, Moers e Mannheim. Para otimizar a eficiência computacional e contornar a limitação de hardware, adota-se uma estratégia de pré-processamento *offline*: os tensores visuais correspondentes às regiões das células são gerados e armazenados em disco previamente, reduzindo a carga de processamento da CPU durante o treinamento. A infraestrutura utilizada consiste em uma estação de trabalho equipada com processador Intel Core i5-10600K, sem o auxílio de aceleradores gráficos (GPU). Esta restrição de

hardware valida a eficiência da arquitetura proposta, demonstrando sua viabilidade em equipamentos de custo acessível.

A definição dos hiperparâmetros segue uma abordagem empírica baseada na estabilidade da convergência. Utiliza-se o otimizador **AdamW** devido à sua capacidade de desacoplar o decaimento de peso (*weight decay*) da otimização do gradiente, favorecendo a generalização. As configurações finais adotadas são:

- **Tamanho do Lote:** Fixado em 64. Experimentos comparativos com lotes de 128 (comuns na literatura) demonstram que o tamanho 64 acelera a convergência em média em 4 épocas e resulta em métricas superiores para os dados utilizados.
- **Taxa de Aprendizado:** Valor inicial de  $1 \times 10^{-3}$  (0.001), com decaimento de peso (*weight decay*) de 0.01.
- **Épocas:** Executado por 70 épocas para garantir a estabilização da função de perda.
- **Particionamento dos Dados:** Particionado em 60% para treinamento, 20% para validação e 20% para testes.

Para a etapa de **Transferência de Aprendizado** (Moers → Frankfurt & Mannheim → Frankfurt), os hiperparâmetros são ajustados para lidar com a escassez de dados do alvo. O tamanho do lote é mantido em 64 e a taxa de aprendizado ajustada para  $7.5 \times 10^{-4}$ , visando preservar o aprendizado adquirido nos domínios de origem enquanto se adapta às especificidades de Frankfurt. A estratégia de particionamento de dados para o ajuste fino (*fine-tuning*) utiliza 30% das amostras locais para ajuste e 70% para teste, simulando um cenário de implantação com poucas medições disponíveis (*drive-test* reduzido). Para a avaliação de desempenho e treinamento, são utilizados a Raiz do Erro Quadrático Médio (*Root Mean Square Error* - RMSE) e o Erro Absoluto Médio (*Mean Absolute Error* - MAE). Embora o RMSE seja a função de perda padrão para otimização (penalizando erros grandes quadraticamente), este trabalho prioriza a análise do MAE para a análise final de qualidade do modelo. Conforme argumentado por Waheed *et al.* [Waheed et al. 2022a], o MAE oferece uma representação mais fiel da precisão do modelo em cenários de telecomunicações, pois atribui peso linear a todas as medições, evitando que *outliers* isolados distorçam a percepção da acurácia geral da cobertura.

## 6. Resultados e Discussão

O processo de treinamento utiliza o conjunto de dados das cidades definidas na Tabela 1 e os hiperparâmetros estabelecidos na Seção 5 para estimar a perda de propagação do percurso entre o transmissor e o receptor do sinal.

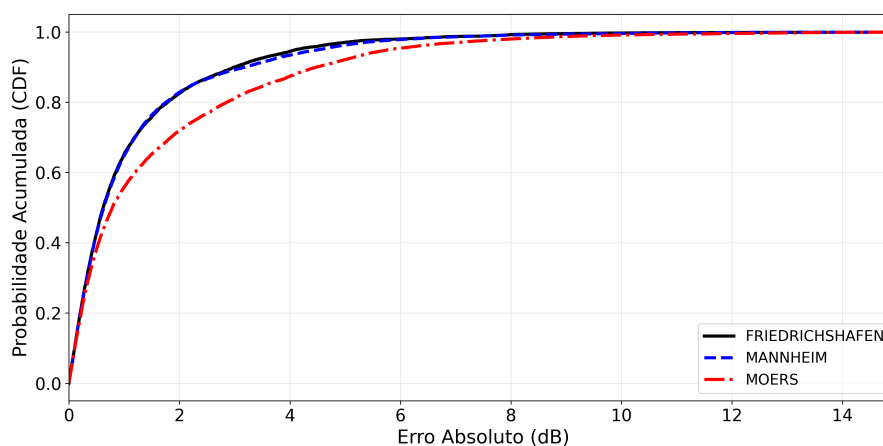
### 6.1. Resultados para as cidades utilizadas no treinamento

A Tabela 4 apresenta os resultados obtidos para o conjunto de testes das cidades utilizadas no treinamento. Para a cidade de Friedrichshafen, na qual o modelo foi treinado nas frequências de 2300 MHz, observam-se as melhores métricas em comparação às demais cidades estudadas, atingindo um RMSE de 1.97 dB e MAE de 1.18 dB. A acurácia superior do modelo para esta cidade segue o comportamento esperado da física de propagação: frequências menores sofrem menos atenuação por difração e absorção, resultando em um comportamento mais previsível do que em bandas milimétricas ou médias. Em comparação, na cidade de Mannheim, o modelo apresenta resultados levemente inferiores, com um acréscimo de aproximadamente 0.03 dB no MAE em relação a Friedrichshafen. Contudo,

**Tabela 4. Resultados obtidos para os conjuntos de teste das cidades avaliadas.**

Cidade	RMSE (dB)	MAE (dB)
Friedrichshafen	1.97	1.18
Mannheim	2.08	1.21
Moers	2.69	1.67

é crucial notar que o modelo para Mannheim opera em frequências quase 3 vezes mais altas. Já para a cidade de Moers, que apresenta métricas mais elevadas (RMSE de 2.67 dB e MAE de 1.67 dB), a diversidade de dados em diferentes frequências auxilia o modelo a generalizar melhor a fenomenologia do problema. Embora isso possa penalizar levemente a acurácia local, resulta em um modelo que compreende melhor a física de propagação global. Por exemplo, considerando uma mesma coordenada absoluta com duas medidas diferentes, uma em 5450 MHz e outra em 5850 MHz, o modelo não apenas memoriza a localização, mas interpreta como o sinal se propaga distintamente em cada frequência naquela região de interesse. A Figura 3 apresenta as Funções de Distribuição Acumulada (CDF) do erro absoluto de previsão para as cidades de Friedrichshafen, Mannheim e Moers. Os resultados evidenciam a robustez do modelo proposto, garantindo que 90% das amostras de teste apresentem erros inferiores a 3.00 dB, 3.19 dB e 4.48 dB, respectivamente.

**Figura 3. Função de Distribuição Acumulada (CDF) do erro absoluto de previsão para as cidades de Friedrichshafen, Mannheim e Moers.**

## 6.2. Transferência de Aprendizado

A transferência de aprendizado é uma técnica de aprendizado de máquina na qual uma rede treinada para um objetivo específico tem sua arquitetura e pesos reutilizados para resolver um problema distinto ou em um novo domínio. Neste processo, realizam-se apenas ajustes finos (*fine-tuning*) no modelo para preservar o conhecimento prévio, refinando os pesos para otimizar o desempenho na nova tarefa. Seguindo essa lógica, a transferência de aprendizado de uma cidade para outra é avaliada em duas etapas. Inicialmente, aplica-se o teste *zero-shot* para verificar a capacidade de generalização do modelo sem nenhum ajuste prévio na cidade alvo. Posteriormente, realiza-se o ajuste fino para quantificar o ganho de acurácia na cidade alvo. A cidade de Frankfurt foi escolhida como alvo devido à sua menor quantidade de dados disponíveis, conforme Tabela 1. A hipótese central é que, se o modelo generaliza bem a física do problema, a campanha de medição necessária para

calibrar o modelo em uma nova cidade pode ser drasticamente reduzida, pressupondo uma amostragem uniforme do espaço. A transferência é realizada a partir dos modelos pré-treinados em Mannheim e Moers. A Tabela 5 explicita os resultados para a transferência, baseando-se nas definições na Seção 6.2. Os resultados confirmam a hipótese de redução

**Tabela 5. Resultados dos experimentos de transferência de aprendizado.**

Cidade de origem	Métrica	Zero-Shot (dB)	Pós-Ajuste (dB)	Melhora (dB)
Moers	RMSE	15.04	2.41	12.63
	MAE	14.41	1.89	12.52
Mannheim	RMSE	24.87	2.65	22.22
	MAE	24.43	2.10	22.33

da campanha de medição. Além disso, observa-se que o modelo treinado com dados de múltiplas frequências (cidade de Moers) atinge uma precisão inicial e final superior em comparação ao modelo treinado em apenas uma banda (Mannheim). Isso corrobora a tese de que a exposição a diferentes frequências durante o treinamento permite que a rede neural aprenda de forma mais robusta a física do ambiente, e não apenas as características geométricas específicas de uma cidade.

## 7. Conclusão

Este trabalho aprofunda o estudo sobre a predição de perda de propagação de sinal, integrando mecanismos de atenção do estado da arte em visão computacional para aprimorar a interpretação física e a acurácia dos modelos neste domínio. A formulação inovadora da imagem de entrada em formato 2.5D, aliada à introdução do mecanismo PSA, demonstra-se importante para o desempenho do sistema. Essa combinação contribui para a melhoria das métricas, tanto em termos de acurácia global, quanto na robustez em cenários de transferência de aprendizado e testes *zero-shot*. Uma observação desta pesquisa é que o treinamento com diversidade espectral (múltiplas frequências), com base nos experimentos feitos, concede ao modelo uma capacidade superior de generalização. Ao ser exposto a diferentes comportamentos de propagação simultaneamente, o modelo evita a memorização de características geográficas específicas, aprendendo de forma mais eficaz a física subjacente ao problema. Como direcionamento para trabalhos futuros, destaca-se a necessidade de um estudo mais rigoroso sobre técnicas avançadas de pré-processamento e filtragem de *outliers*. Tal refinamento impede que a rede neural sobreajuste ruídos de medição ou incertezas estocásticas, garantindo que o aprendizado foque nas características determinísticas do ambiente local. Adicionalmente, é possível avaliar otimizações na arquitetura da rede neural visando a eficiência computacional, buscando modelos mais enxutos que facilitem o treinamento sem perda de desempenho. Assim como, pretende-se expandir o conjunto de dados para abranger faixas de frequência ainda mais amplas e melhorar o estudo de transferência de aprendizado, consolidando a hipótese de que a variabilidade das frequências é a chave para a criação de preditores de canal universalmente aplicáveis. Por fim, ampliar a pesquisa com um estudo de ablação relacionado à adição e modificações da arquitetura em comparação à de referência.

## Agradecimentos

Este trabalho foi realizado com apoio da Anatel, por meio do TED sob o processo SEI nº 53500.060216/2024-35 e com apoio da Coordenação de Aperfeiçoamento de

Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001. O trabalho também foi financiado pelo CNPq (408255/2023-4, 407304/2025-8, 308917/2025-1), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), auxílio E-26/204.562/2024, e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), auxílios no. 2023/00673-7 e 2023/00811-0.

## Referências

- Cavalcanti, B., Cavalcante, G., and Mendonça, L. (2025). Performance analysis of artificial neural networks for predicting propagation losses in suburban environments for 4g lte and 5g networks. *Revista Principia*, 62.
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., and Raskar, R. (2018). Deepglobe 2018: A challenge to parse the earth through satellite images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- Lee, J.-Y., Kang, M. Y., and Kim, S.-C. (2019). Path loss exponent prediction for outdoor millimeter wave channels through deep learning. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–5.
- Park, D., Ryu, S., Lee, S., Kang, N., Kim, S., Kim, K., Choi, D., and Ahn, S. (2024). 5g base station electromagnetic field strength estimation method in complex hotspot area using deep learning. In *2024 IEEE International Symposium on Electromagnetic Compatibility, Signal & Power Integrity (EMC+SIPI)*, pages 551–554.
- Sani, U. S., Malik, O. A., and Lai, D. T. C. (2022). Improving path loss prediction using environmental feature extraction from satellite images: Hand-crafted vs. convolutional neural network. *Applied Sciences*, 12:7685.
- Vasudevan, M. and Yuksel, M. (2024). Machine learning for radio propagation modeling: A comprehensive survey. *IEEE Open Journal of the Communications Society*.
- Waheed, M. T., Fahmy, Y., and Khattab, A. (2022a). Deepchannel: Robust multi-modal outdoor channel model prediction in lte networks using deep learning. *IEEE Access*, 10:79289–79300.
- Waheed, M. T., Khattab, A., and Fahmy, Y. (2022b). Highly accurate multi-modal lte channel prediction via semantic segmentation of satellite images. In *2022 10th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*, pages 90–93.
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., and Ding, G. (2024). Yolov10: Real-time end-to-end object detection.
- Yeaser, K. M. A. and Hassan, K. M. A. (2025). A comprehensive study on path loss estimation using deep hybrid learning in 5G networks. *Journal of Telecommunications and Information Technology*, pages 86–94.
- Zeng, S., Ji, Y., Chen, W., Yan, L., and Zhao, X. (2024). Mobile network coverage prediction using multi-modal model based on deep neural networks and semantic segmentation. *Sensors*, 24(16).