

# Identificação de composições da paisagem urbana: uma abordagem de *deep learning*

Ana Luiza Favarão Leão<sup>a</sup> , Hugo Queiroz Abonizio<sup>b</sup> , Sylvio Barbon Júnior<sup>c</sup> ,  
Milena Kanashiro<sup>d</sup> 

<sup>a</sup> Universidade Estadual de Londrina, Departamento de Arquitetura e Urbanismo, Programa de Pós-Graduação em Arquitetura e Urbanismo, Londrina, PR, Brasil. E-mail: analuiza.favarao@uel.br

<sup>b</sup> Universidade Estadual de Londrina, Departamento de Computação, Programa de Pós-Graduação em Ciência da Computação, Londrina, PR, Brasil. E-mail: hugo.abonizio@uel.br

<sup>c</sup> Universidade Estadual de Londrina, Departamento de Computação, Programa de Pós-Graduação em Ciência da Computação, Londrina, PR, Brasil. E-mail: barbon@uel.br

<sup>d</sup> Universidade Estadual de Londrina, Departamento de Arquitetura e Urbanismo, Programa de Pós-Graduação em Arquitetura e Urbanismo, Londrina, PR, Brasil. E-mail: milena@uel.br

<https://doi.org/10.47235/rmu.v8i1.140>

Submetido em 8 de março de 2020. Aceito em 21 de maio de 2020.

---

**Resumo.** *A composição do ambiente pode exercer impactos sobre seus usuários, no entanto, esta relação permanece incerta até que as composições da paisagem urbana e suas qualidades espaciais possam ser analisadas empiricamente. Imagens obtidas através do Google Street View (GSV) possibilitam um grande volume de dados para avaliação automatizada das características ambientais. Técnicas de deep learning têm avançado na identificação de elementos compositivos do ambiente construído. Neste sentido, este estudo busca investigar e testar um procedimento de identificação da configuração e composição da paisagem urbana, por meio da classificação de imagens obtidas pelo GSV. A partir de um banco de imagens de três bairros de Londrina-PR, um modelo de deep learning para classificação de imagens foi proposto. O modelo obteve um bom desempenho, atribuindo corretamente 87,6% das amostras dos respectivos bairros do estudo de caso. Características compositivas foram empiricamente identificadas, considerando a distribuição das amostras no espaço de busca obtido. O modelo proposto contribui na definição de recortes espaciais bem como na mensuração de qualidades ambientais, otimizando coletas de dados, ampliando amostras e conferindo objetividade aos resultados. Esta abordagem contribui na expansão das escalas analíticas da cidade, identificando padrões compositivos e relacionais para o entendimento de elementos influentes no comportamento humano.*

**Palavras-chave.** *morfologia urbana, ambiente construído, aprendizado profundo, classificação de imagens, Google Street View.*

---

## Introdução

A cidade como um artefato antrópico, com diferentes configurações, reflete a complexidade das relações socioespaciais. Assim, composição e forma da cidade, resultado de características naturais e construídas, são parâmetros importantes em

análises voltadas ao Planejamento e Projeto Urbano. A composição do ambiente construído tem sido associada a atividade física realizada pelos residentes (Sallis et al., 2015), sua satisfação com o ambiente (Lee et al., 2017), segurança contra o crime (Kamalipour, Faizi e Memarian, 2014) e até a felicidade dos indivíduos (Kent, Ma e

Mulley, 2017; Seresinhe *et al.*, 2019). Neste sentido o design urbano pode influenciar as escolhas e o comportamento das pessoas. No entanto, esta relação permanece incerta até que as composições da paisagem urbana e suas qualidades espaciais possam ser definidas, quantificadas, medidas e testadas empiricamente (Ewing e Handy, 2009).

Dentre as diversas abordagens teórico-metodológicas de análise do ambiente urbano, desde leituras históricas, morfológicas, de apropriações, de percepções, ou de temporalidades, uma das estratégias é a definição de recortes espaciais, a partir do reconhecimento de padrões. Por exemplo, a delimitação de Unidades de Paisagem permite a identificação dos atributos responsáveis pela dinâmica da paisagem (Amorim e Oliveira, 2008); a definição de setores/distritos/bairros define limites para agregação de dados para a caminhabilidade (Gehrke e Wang, 2020); a concepção de unidades homogêneas urbanas para o agrupamento de áreas com as mesmas características ambientais e/ou físico-espaciais para o estabelecimento de classes de usos urbanos (Medeiros e Grigio, 2019). Tais processos buscam a definição de padrões espaciais, a partir de parâmetros pré-definidos. Para esse procedimento, análises de características individuais e posteriores sobreposições de informações para a delimitação de áreas têm sido realizadas por meio de levantamentos de dados secundários, de campo e de definição de critérios.

Nos últimos anos, imagens urbanas de produtos online como *Google Street View* (GSV) têm se tornado cada vez mais disponíveis, retratando o ponto de visão do usuário (Middel *et al.*, 2019). Por meio do uso destas ferramentas, se torna possível testar abordagens *de big data* e de aprendizado de máquina para estudos de caracterização visual de elementos tipológicos (Doersch *et al.*, 2012; Yin e Wang, 2016; Liu *et al.*, 2017; Moosavi, 2017; Zhang *et al.*, 2017; Shen *et al.*, 2018), a mensuração de qualidades urbanas que possivelmente influenciam o comportamento e bem-estar na escala da rua (Yin *et al.*, 2015; Yin e Wang, 2016; Liu *et al.*, 2017), assim como o uso de imagens de alta

resolução para os estudos de morfologia urbana (Moosavi, 2017; Shen *et al.*, 2018; Zhang *et al.*, 2017). Este tipo de procedimento pode ser uma alternativa mais eficiente (Ben-Joseph *et al.*, 2015) se comparado à levantamentos ou auditorias em campo que além de possivelmente sujeitos a subjetividade, são mais onerosos e requerem deslocamentos em locais nem sempre seguros para o pesquisador (Badland *et al.*, 2010).

Por outro lado, os métodos de aquisição de imagens urbanas como o GSV possibilitam a obtenção de um grande volume de dados para a realização das análises. Desta forma, a avaliação automatizada de características compositivas do ambiente demonstra potencial para incorporação na área. Neste contexto, abordagens que utilizam aprendizado de máquina têm ganhado espaço na literatura. Tais métodos objetivos de obtenção de dados do espaço urbano diminuem a necessidade de despende tempo em levantamentos *in loco*, ainda possibilitando um maior volume de dados (Zhang *et al.*, 2018).

Entretanto, o aprendizado de máquina tradicional limita-se à necessidade de extração de características previamente definidas que descrevam o problema e à sua capacidade de lidar com dados de alta dimensionalidade (e.g. pixels de uma imagem obtida através do GSV), resultando no problema conhecido como Maldição da Dimensionalidade (Poggio *et al.*, 2017). Em contrapartida, métodos de *deep learning* apresentam múltiplas camadas de processamento e representação de dados, por meio da composição de módulos não lineares que transformam a representação inicial (e.g. pixels de uma imagem), em elementos de um maior nível de abstração (Lecun, Bengio e Hinton, 2015). Técnicas de *deep learning*, como redes neurais convolucionais, têm avançado com aplicações de visão computacional em muitos domínios de pesquisa, inclusive na identificação de elementos compositivos do ambiente construído como a estética (Tan *et al.*, 2017) e paisagem (Zhou *et al.*, 2016).

Considerando tais aspectos metodológicos e conceituais, esta pesquisa apresenta como

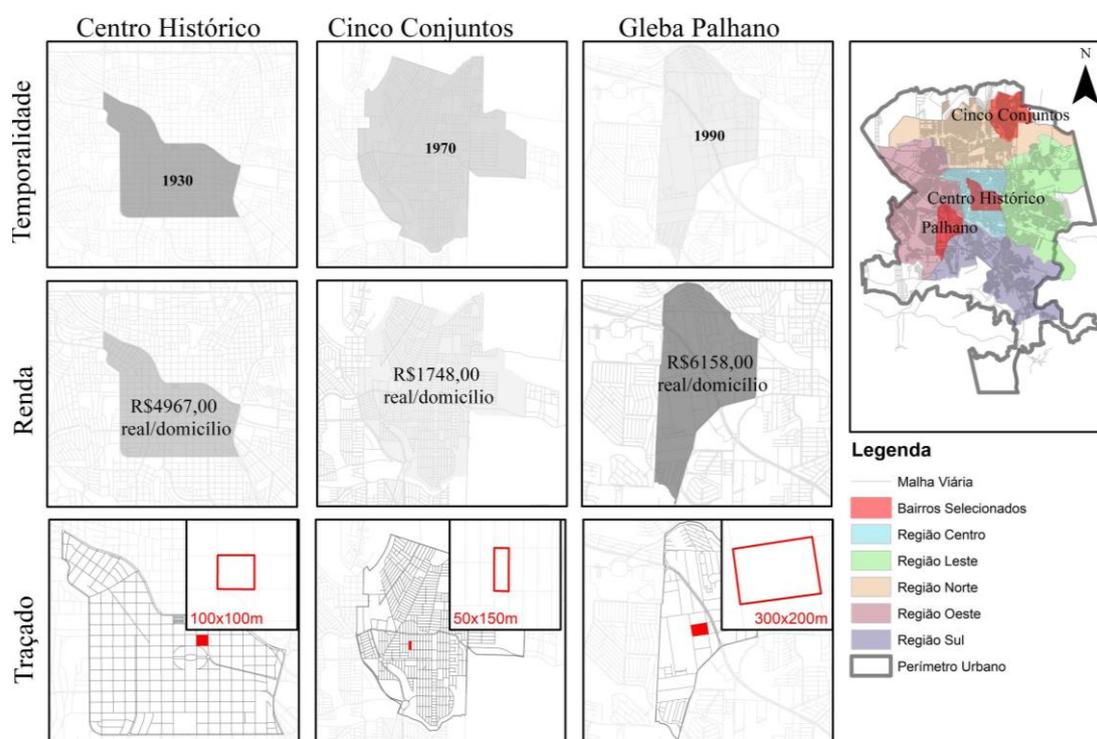
objetivo geral, testar um procedimento de identificação da configuração e composição da paisagem urbana, por meio de uma abordagem de classificação de imagens na aplicação de um método de *deep learning*. A partir da hipótese de que as imagens avaliadas, por meio de um modelo de classificação, permitirão o reconhecimento de padrões do ambiente construído em imagens obtidas através do GSV. Este trabalho explora e discute a aplicação de modelos de *deep learning* como ferramenta para a análise da paisagem urbana. Em suma, busca-se agrupar composições similares geo-informativas (Doersch et al., 2012) da paisagem. A partir de um banco de imagens, características compositivas foram identificadas, considerando a distribuição das amostras no espaço de busca construído através de um modelo de classificação.

Os resultados da presente pesquisa contribuem no fomento de pesquisas com aplicações de *deep learning* para a compreensão de leituras de paisagem urbana

de forma objetiva e baseada em evidências. O modelo criado obteve uma acurácia de 87,6%, indicando uma boa capacidade de generalização. Este resultado demonstra um bom desempenho do modelo, indicando que características compositivas que variam entre os bairros foram corretamente abstraídas, aprendidas e reconhecidas. /um prisma de caráter subjetivo e limitado a pequenas amostras (Yin e Wang, 2016).

## Métodos

A estratégia de pesquisa adotada foi o estudo de caso, visto que o fenômeno analisado é contemporâneo e contextual, portanto indissociável da realidade (Yin, 2001). Como caso, três setores definidos pelo município foram adotados como recorte na cidade de Londrina-PR, considerando, principalmente, suas diferenças socioeconômicas, temporalidades de desenvolvimento, traçado urbano e localização geográfica: Centro Histórico, Cinco Conjuntos e Gleba Palhano (Figura 01).



**Figura 1.** Bairros selecionados (fonte: elaborado pelos autores, 2020).

Londrina apresenta uma população estimada, em 2019, de 569.733 hab. (IBGE, 2018). A cidade foi construída, a partir de um pré-projeto em 1932, como parte do processo de

colonização da Companhia de Terras Norte do Paraná, no contexto histórico da produção cafeeira. Os anos seguintes foram crescentemente marcados por um

crescimento urbano ao redor de um núcleo planejado (Töws, Mendes e Vercezi, 2010), que configura hoje o bairro chamado de Centro Histórico. Características de ortogonalidade da malha, intenso uso misto e adensamento residencial vertical, marcam a valorização do núcleo inicial da cidade.

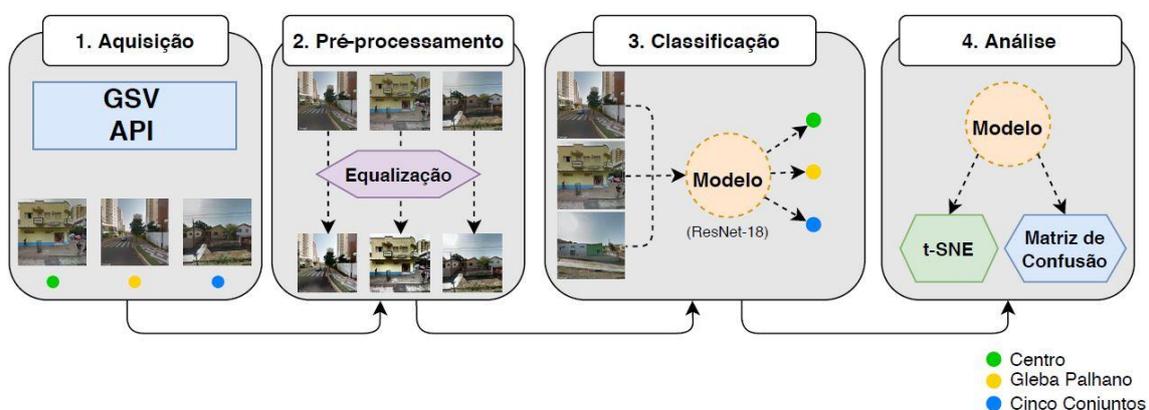
A Gleba Palhano, um setor localizado na Zona Sudoeste de Londrina, iniciou sua consolidação a partir dos anos 90. Inúmeros fatores condicionaram o desenvolvimento desta área, mas principalmente o apelo ao mercado imobiliário pela localização privilegiada dentro da malha urbana, sendo até hoje objeto de grande valorização fundiária. Atualmente é a região mais valorizada da cidade, superando o Centro Histórico (Oura, 2006). Caracterizada por uma intensa ocupação verticalizada, o padrão de condomínios alocados em grandes lotes é direcionado para a população de alta renda.

O desenvolvimento da Zona Sudoeste de Londrina se contrapõe ao da Zona Norte, caracterizada, sobretudo pelos empreendimentos de habitação de interesse social, para a classe baixa e focos de favelamento (Töws, Mendes e Vercezi, 2010). O Cinco Conjuntos, a terceira área selecionada, situa-se na Zona Norte e tem sua origem em projetos habitacionais da década de 1970. Este empreendimento foi mediado pela logística de construções de casas da Companhia de Habitação – COHAB/LDA, responsável por subsidiar verba para construção de habitações de baixa renda por toda periferia de Londrina (Beidack e Fresca,

2011). A partir da década de 1990, com um adensamento populacional consolidado, uma nova dinâmica, a avenida Saul Elkind estabeleceu-se como centralidade com atividades de comércio e prestação de serviços.

Para criação do modelo de identificação dos bairros foi utilizada uma abordagem de aprendizado de máquina supervisionado, no qual um conjunto de amostras rotuladas com a classe correspondente é utilizado para modelar a distribuição condicional das classes que compõem o problema (Bishop, 2006). Nesta abordagem um modelo é induzido por pares de entradas e saídas que são utilizados no aprendizado de padrões presente nos dados. Neste trabalho tem-se como entrada para o modelo uma imagem da paisagem urbana e como saída o bairro a qual a imagem pertence.

O desenvolvimento do modelo e análise dos resultados foram divididos em quatro etapas: (1) a coleta de dados, na qual foi realizada a aquisição das imagens disponíveis no GSV; (2) o pré-processamento, em que as imagens passam por uma fase de tratamento que visa melhorar a qualidade dos dados; e (3) a etapa de indução do modelo, para a realização do treinamento utilizando as amostras que foram pré-processadas e o algoritmo de aprendizado de máquina; e, por fim, (4) a análise dos resultados por meio de técnicas de interpretação do modelo. A Figura 02 ilustra as etapas do processo, detalhadas na sequência.



**Figura 2.** Etapas do desenvolvimento metodológico da pesquisa (fonte: elaborado pelos autores, 2020).

### Aquisição

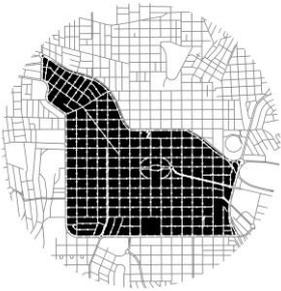
Uma amostragem de pontos distribuídos de 100 em 100 metros pela rede viária foi realizada por meio do software ArcGis 10.6. O toolbox de Gerenciamento de dados foi utilizado, com a ferramenta Generate Points Along Lines, que aloca pontos ao longo de linhas em intervalos fixos. Considerando dados de centro de via disponibilizados pela Prefeitura de Londrina (Prefeitura do Município de Londrina, 2020), a medida adotada foi utilizada por representar a dimensão aproximada das quadras do plano inicial da cidade (Yamaki, 2017, p. 64).

Por meio da API (Application Programming Interface) do GSV, imagens foram obtidas a partir da localização geográfica dos pontos amostrais. Os seguintes parâmetros foram utilizados para as requisições: câmera orientada ao norte, rotação vertical neutra em relação ao solo e definição de campo de visão

de 90 graus. As amostras que resultaram em um erro de aquisição pela API, devido à falta de imagens para a dada coordenada, foram descartadas.

Um total de 2.017 amostras foi obtido, conformando a base de imagens utilizada nesta pesquisa (Tabela 01). A amostragem totalizou n=554 pontos no setor do Centro Histórico, n=368 pontos na Gleba Palhano e n=1095 no Cinco Conjuntos. Houve um desequilíbrio do número amostral por setor, em virtude, principalmente, da quantidade de vias existentes. Aponta-se que o setor do Cinco Conjuntos além de possuir a maior dimensão em área, possui uma rede viária mais extensa (Tabela 01) em função de um macro parcelamento marcado por quadras retangulares com lotes de dimensões mínimas.

**Tabela 1.** Características da malha e amostragem (fonte: elaborado pelos autores, 2020).

	Cinco Conjuntos	Centro Histórico	Gleba Palhano
Setor			
Extensão	129,76 km	49,68 km	61,16 km
Rede Viária	6,80km <sup>2</sup>	4,26km <sup>2</sup>	3,26km <sup>2</sup>
Amostra	1095	554	368

### Pré-Processamento

A partir das amostras coletadas, o passo seguinte foi o pré-processamento das imagens, visando melhorar a qualidade das predições produzidas pelo modelo e evitando o sobreajuste em características pouco informativas. Considerando que as imagens foram adquiridas pelo GSV em diferentes horários do dia e em diferentes condições

climáticas, foi realizado um tratamento para evitar que características como a iluminação e a saturação fossem evidenciadas pelo modelo e reconhecidas como informações relacionadas ao bairro.

Para remover as diferenças de iluminação, foi utilizado um espaço de cor onde, diferente do padrão RGB, a iluminação é isolada dos outros componentes da imagem. Sendo

assim, as imagens foram convertidas para o espaço de cor YCbCr, utilizando a biblioteca OpenCV<sup>1</sup>. YCbCr é um espaço de cor em que o canal Y representa a luminância, o canal Cb a croma azul e Cr a croma vermelha (Gowda e Yuan, 2019). Dessa forma, o histograma do canal de luminância foi utilizado para equalizar as imagens. Criou-se assim, uma distribuição mais ampla e uniforme dos valores de intensidade do canal Y, atenuando as diferenças causadas pelos contrastes da iluminação.

### Classificação

Após o pré-processamento, com o conjunto de imagens adequadas para modelagem, o passo seguinte foi a criação do modelo de classificação de imagens. Com o intuito de avaliar o desempenho do modelo sem nenhum viés indutivo, as imagens foram divididas em um conjunto de treino e um conjunto de teste, onde o primeiro é utilizado para a indução do modelo e o segundo é utilizado para avaliar o desempenho do modelo induzido. Sendo assim, o conjunto de teste serve como parâmetro para quantificar a capacidade de generalização do modelo em novas amostras, ou seja, amostras não utilizadas durante a fase de treinamento (Bishop, 2006). Dessa forma, as amostras foram separadas aleatoriamente e de maneira estratificada em relação a classe em um conjunto contendo 80% das imagens utilizado para o treinamento do modelo e os 20% restantes foram utilizados para medir o desempenho.

Em seguida, para a criação do modelo de classificação de imagens, foi selecionada a arquitetura ResNet (He e Sun, 2016), uma Rede Neural Convolutiva que utiliza blocos residuais e conexões de atalho para evitar o problema da dissipação do gradiente (Glorot e Bengio, 2010). Essa arquitetura se mostrou eficaz em uma grande variedade de problemas, incluindo aplicações em análise de imagens médicas (Litjens et al., 2017), agricultura (Kamilaris e Prenafeta-Boldú, 2018), descrição automática de cenas (Anderson et al., 2018) e identificação de pedestres (Fan et al., 2018). Sendo assim, a arquitetura ResNet se mostra robusta e

adequada para diversas aplicações em análise de imagem.

Foi utilizada uma rede ResNet com 18 camadas (ResNet-18) previamente treinada com o banco de dados ImageNet (Deng et al., 2009). Tal abordagem tem como objetivo utilizar os padrões aprendidos em grandes conjuntos de dados – como o ImageNet, que contém milhões de amostras – e transferir este aprendizado para alavancar a identificação de características com uma quantidade menor de amostras. Estes padrões vão desde características simples como linhas e bordas até características mais complexas como formas poligonais e texturas em camadas mais profundas da rede.

A partir da rede inicializada com os pesos previamente treinados, foi realizada a sintonia fina do modelo ao longo de 30 épocas usando o algoritmo Adam (*Adaptive Moment Estimation*) (Kingma e Ba, 2014) para otimizar os parâmetros. Dessa forma, os parâmetros aprendidos no treinamento prévio são ajustados no conjunto de dados da base investigada, aprendendo a classificar as imagens em seu respectivo bairro.

Uma vez que existe um desbalanço entre as classes, isto é, quantidades diferentes de amostras em cada classe (i.e. bairros), sendo uma delas significativamente maior que as outras, foram utilizados diferentes pesos entre as classes para compensar a representatividade de cada classe no conjunto total. Também foram aplicadas técnicas tradicionais de *data augmentation*, como rotação horizontal, que dobra a variação das imagens espelhando o eixo horizontal, e recorte aleatório, que cria diversas variações da imagem mantendo o mesmo centro (Takahashi, Matsubara e Uehara, 2019; Shorten e Khoshgoftaar, 2019). Tais procedimentos foram implementados através da biblioteca de aprendizado de máquina *PyTorch*<sup>2</sup> e seus parâmetros padrão.

### Análise

Com o objetivo de compreender os padrões reconhecidos pelo modelo e interpretar as características modeladas, foi realizada uma visualização em duas dimensões das amostras presentes no conjunto de teste

utilizando uma técnica de redução de dimensionalidade. A técnica escolhida foi t-SNE (*t-Distributed Stochastic Neighbor Embedding*), um método não linear de visualização de dados com uma alta dimensionalidade que atribui a cada ponto de um espaço N-dimensional uma localização em um novo espaço 2D ou 3D (Maaten e Hinton, 2008).

O algoritmo funciona criando uma distribuição de probabilidade Gaussiana baseada na similaridade entre cada par de amostras no espaço original. Assim, essa distribuição é recriada utilizando a distribuição t-Student no novo espaço de baixa dimensionalidade. A otimização da distribuição na nova projeção é feita através de gradiente descendente na divergência de Kullback-Leibler entre as duas distribuições. Diferente de outras técnicas de redução de dimensionalidade que levam em consideração apenas relações lineares no espaço de características, como a Análise de Componentes Principais (Borges, 2010), a t-SNE mapeia também relações não lineares complexas da estrutura local e global dos dados.

Para visualização das amostras no espaço de características do modelo treinado, foi extraída a representação resultante da última camada de convolução da rede, antes de ser passada para a camada de classificação. Dessa forma, são extraídos vetores com 512 valores para cada imagem, representando as características aprendidas pelo modelo para classificação das imagens. A partir dos vetores de características das imagens, as amostras foram então projetadas em duas dimensões utilizando t-SNE (Chan et al., 2018).

Além da visualização das amostras no espaço de características, também foi utilizada a matriz de confusão (Powers, 2011), a qual apresenta a contagem de amostras em relação ao seu rótulo real e a predição feita pelo modelo de classificação. Dessa forma, é possível analisar o desempenho do modelo em relação a cada classe e as relações de similaridade aprendidas entre elas, na

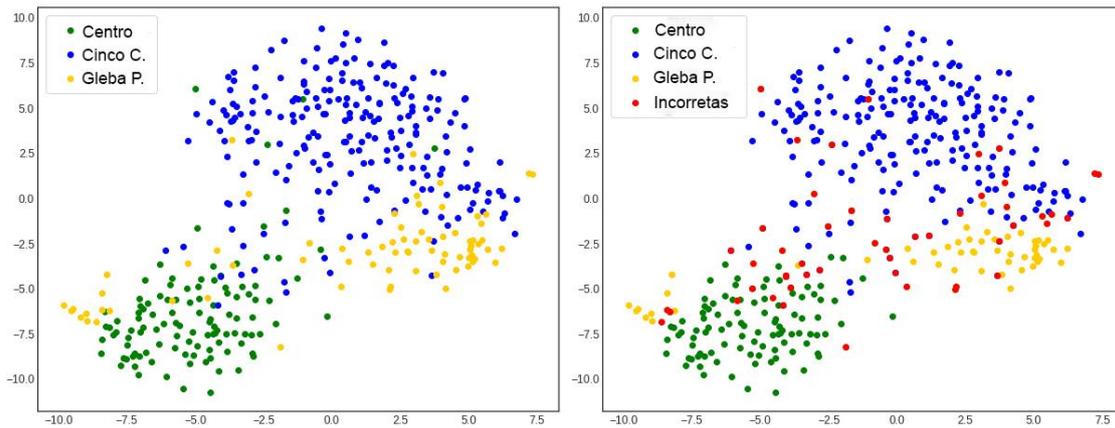
comparação entre o rótulo real das amostras e o rótulo obtido pela classificação.

## Resultados

Para visualização das amostras no espaço de características aprendidas pelo modelo, foi utilizada a técnica t-SNE. A Figura 03 apresenta dois gráficos com a distribuição das amostras do conjunto de teste, diferenciando o bairro a qual pertence a amostra pela cor e destacando as amostras que foram erroneamente classificadas. Assim, a partir desta projeção foi possível identificar padrões de agrupamentos das amostras na estrutura do espaço de características aprendidas.

O gráfico à esquerda na Figura 03(a) apresenta duas grandes aglomerações dos bairros Centro e Cinco Conjuntos, indicando uma clara divisão das características detectadas nas imagens dos dois bairros. A Gleba Palhano, em uma posição intermediária, divide-se em dois grupos, um mais próximo das amostras do bairro Cinco Conjuntos e outro do Centro. Essa divisão em dois subgrupos decorreu de características no bairro Gleba Palhano se assemelharem ao Cinco Conjuntos e ao Centro. Parte do bairro possui uma estrutura recente verticalizada - se assemelhando às características encontradas no Centro - enquanto em outras partes ainda são encontrados muitos terrenos vazios e áreas em desenvolvimento.

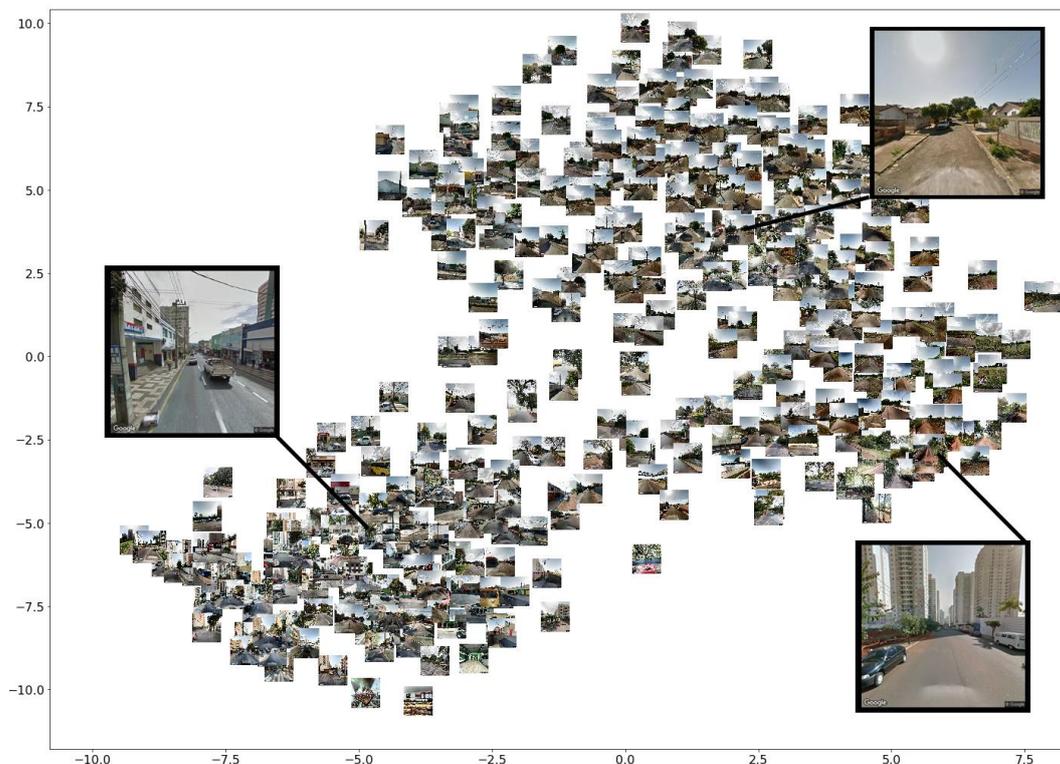
O gráfico à direita na Figura 03(b) identifica as amostras que tiveram o rótulo atribuído pelo modelo diferente do bairro a qual ela pertence. É importante destacar que existe uma área entre as aglomerações das classes que é mais 'nebulosa', não havendo uma distinção linear entre elas. Desta forma, é possível identificar amostras do bairro Centro inseridas no grupo do bairro Cinco Conjuntos, acontecendo o mesmo para as outras combinações de classe. Tal fato foi decorrência das características similares da paisagem presentes nestas áreas. Portanto, justamente as amostras nas áreas de transição entre as classes, locus de maior perturbação, foram aquelas que o modelo erroneamente predisse o bairro.



**Figura 3.** Espaço de características - Projeção em duas dimensões das amostras utilizando t-SNE; (a) à esquerda: formação de aglomerações de amostras da mesma classe; (b) à direita: destaque das amostras classificadas erroneamente pelo modelo (fonte: elaborado pelos autores, 2020).

Além da visualização da divisão das classes das amostras no espaço de características, a projeção em duas dimensões das amostras com as imagens originais em cada ponto do espaço bidimensional foi gerada (Figura 04). Desta forma, é possível observar a similaridade, das amostras em conjuntos próximos no espaço, em comparação com as amostras em conjuntos mais distantes. Os agrupamentos reconhecidos na Figura 03 são

facilmente identificados na Figura 04, evidenciando a diferença das características na composição da paisagem do Cinco Conjuntos, na parte superior, e do Centro, na parte inferior. Os dois grupos separados de amostras da Gleba Palhano também são identificáveis nas áreas próximas ao Cinco Conjuntos, na parte direita do gráfico, e do Centro, no canto inferior esquerdo.



**Figura 4.** Projeção em duas dimensões das amostras utilizando t-SNE exibindo as imagens coletadas em cada ponto do espaço bidimensional (fonte: elaborado pelos autores, 2020).

Foi realizada a matriz de confusão comparando o rótulo real das amostras, ou seja, o bairro ao qual a amostra originalmente pertence, e o rótulo predito pelo modelo de classificação no conjunto de teste (Tabela 02). Foi possível entender quais foram os principais erros do modelo em relação a cada classe do problema. Para analisar o número de amostras corretas e incorretas deve ser considerado o desequilíbrio de amostras entre as classes.

O bairro em que o modelo de classificação obteve o melhor desempenho foi o do bairro Cinco Conjuntos, onde apenas 6% das amostras do conjunto de testes foram classificadas erroneamente. Tal fato pode ser explicado por duas questões: esta classe é a que apresenta a maior quantidade de amostras no problema, o que representa uma grande variabilidade de composições para serem aprendidas pelo modelo no conjunto de treino e generalizadas para o conjunto de teste; outra explicação para o melhor desempenho nesta classe é a homogeneidade da paisagem, fruto de empreendimentos de habitação de interesse social predominantes no local

O Centro obteve o desempenho mediano, com equívoco em 12% das amostras do conjunto de teste. Apesar de apresentar uma menor dispersão no espaço de características, comparativamente às amostras do Cinco Conjuntos, o modelo identificou amostras de outras classes com características morfológicas similares. Isso ocorre provavelmente em função de aspectos temporais que caracterizam o local com elementos inerentes a sedimentação do espaço central da cidade.

O pior desempenho foi da Gleba Palhano, com 31% das amostras classificadas equivocadamente. A classe está aglomerada

principalmente em duas regiões, entretanto, existem amostras dispersas na área intermediária entre as classes. Essa ocorrência pode estar associada ao menor número de amostras da classe. Este fato dificulta o aprendizado das características durante a fase de treinamento em função da redução de variabilidade no espaço. A Figura 03 mostra que foram os pontos de maior dificuldade de distinção entre as classes.

A Figura 05 exemplifica situações nas quais o modelo erroneamente classificou as amostras. Em relação a predições do bairro Centro Histórico, imagens incorretamente preditas como pertencentes ao bairro Cinco Conjuntos podem ser retratadas pelo exemplo que contempla características de uma paisagem residencial, no caso com elementos estéticos de tipologias residenciais mais antigas da cidade como a platibanda. Ainda, exemplifica-se uma predição do Centro Histórico como pertencente à Gleba Palhano através de uma imagem que retrata uma paisagem mista, com elementos de verticalização e unidades de apenas 1 pavimento. O bairro Cinco Conjuntos apresentou amostras erroneamente preditas como sendo do Centro Histórico principalmente em imagens que retratam tipologias tipicamente comerciais, bastante típicas do centro da cidade. Em contrapartida, amostras representativas do bairro Cinco Conjuntos foram interpretadas como sendo da Gleba Palhano, em muitos casos, quando vazios urbanos eram retratados. A Gleba Palhano foi erroneamente classificada como Centro Histórico em imagens que retratam principalmente tipologias comerciais e fachadas ativas. Por outro lado, foi predita como Cinco Conjuntos em imagens de vazios urbanos.

**Tabela 2.** Matriz de confusão (fonte: elaborado pelos autores, 2020).

		Rótulo Predito		
		Centro	Cinco Conjuntos	Gleba Palhano
Rótulo Real	Centro	98	8	5
	Cinco Conjuntos	6	205	8
	Gleba Palhano	10	13	51

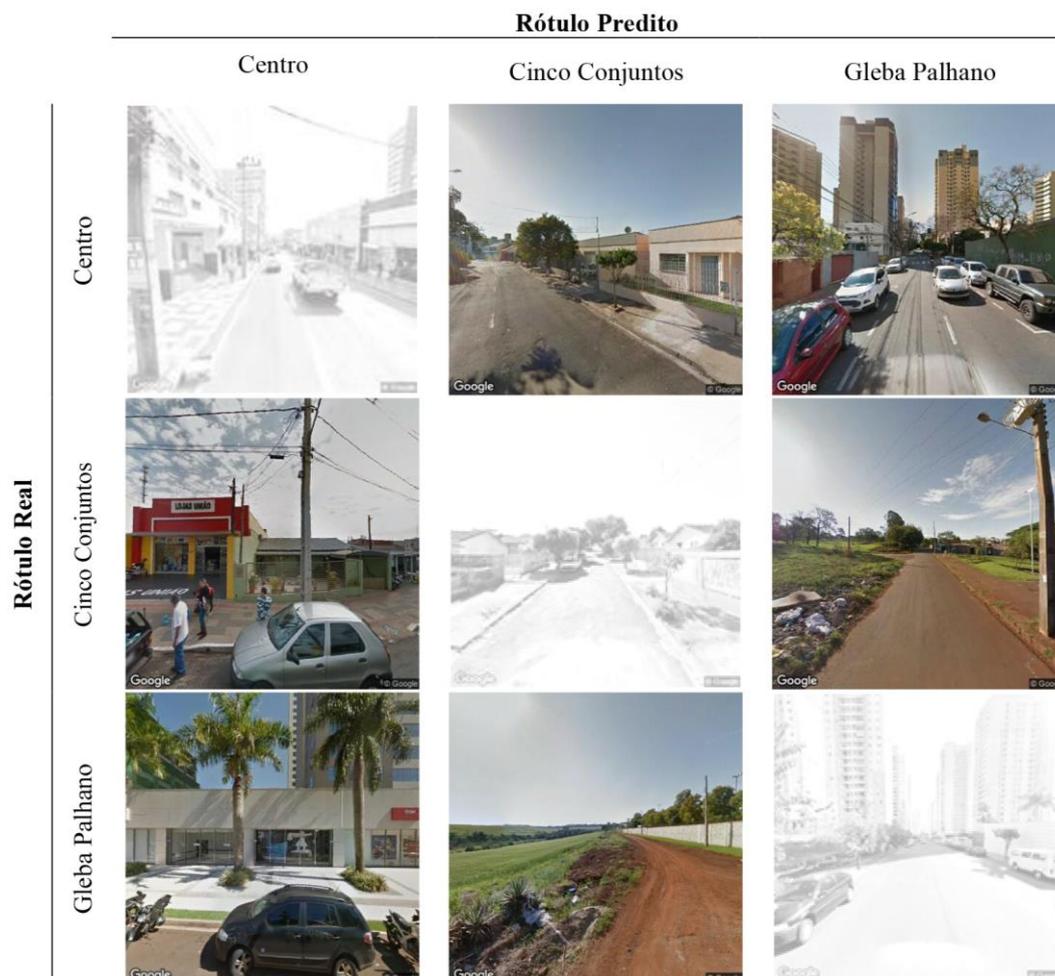


Figura 5. Imagens incorretamente classificadas (fonte: elaborado pelos autores, 2020).

### Discussões

Os resultados demonstram a logicidade no uso do aprendizado de máquinas para a identificação de características da paisagem urbana. O modelo apresentou uma maior capacidade preditiva ao extrair as principais particularidades do setor da Zona Norte da cidade - áreas de empreendimentos de habitação de interesse social - de uma homogeneidade de edificações horizontais. Tais características aumentam a proporção de céu nas imagens, e pela localização periférica a linha do horizonte se dá de forma constante.

No contraponto, no centro da cidade, as áreas verticalizadas de configuração mais compacta e padrões de edificações comerciais e de serviços e, conseqüentemente com a presença de fachadas ativas, foram identificadas de maneira coerente. Elementos como a quantidade de carros, presença de

estacionamentos nas vias e sinalização intensa de trânsito parece ter contribuído para a correta identificação.

Pode-se pontuar que as amostras que representam edifícios comerciais do centro e da zona norte foram diferenciadas, na sua maioria, corretamente inseridas em seus contextos, indicando padrões tipológicos de diferentes categorias (vicinal ou não) e uma leitura relacional da composição da paisagem entre os setores.

A vegetação parece ser outro elemento importante de identificação da paisagem urbana no uso da visão computacional. Tem-se como suposição que fatores relacionados não apenas a existência, mas também os tipos de vegetação, a maturação e a manutenção das áreas vegetadas foram indicativos da associação correta de imagens de vegetação nos respectivos setores analisados.

Classificações incorretas foram detectadas principalmente no setor Sudoeste da cidade, a Gleba Palhano. Conjectura-se que este resultado se deu em função da recência da região, ainda em consolidação, assim compartilhando características relevantes com os outros dois bairros. A existência de vazios e antigos remanescentes de edificações isoladas encontradas na Gleba Palhano são características também da Zona Norte. Ainda, a recente verticalização é uma similaridade com a área Central. Portanto, entende-se que as imprecisões decorreram pelo processo de consolidação do setor.

A passagem do tempo e a manutenção dos elementos físicos do ambiente urbano são indicativos nem sempre objetivos em avaliações da paisagem, pois apresentam um fator de leitura vinculado à percepção/cognição dos profissionais. Porém, na aplicação da metodologia de classificação de imagens, conjectura-se que a temporalidade foi interpretada no processamento das imagens, tanto na vegetação quanto no desgaste aparente dos materiais das edificações.

Leituras clássicas do ambiente, como a de Christopher Alexander (1979), propõem que elementos como prédios, muros, ruas e cercas, formam padrões interrelacionados. Entretanto a identificação destes padrões é marcada pela subjetividade. A identificação da configuração e composição da paisagem através da metodologia utilizada aponta para o avanço no uso de procedimento de aprendizado de máquina, por meio da classificação de imagens como estratégia inovadora paralela às leituras urbanas tradicionais. A visão computacional implementada com o uso de *deep learning* permitirá expandir as escalas analíticas da cidade, identificando seus padrões compositivos e relacionais.

## Conclusões

Por meio da aplicação do procedimento de identificação da configuração e composição da paisagem urbana, no uso de um método de *deep learning* para classificação de imagens, demonstrou-se o potencial desta abordagem para identificação de características do ambiente construído. Apresentamos uma

abordagem efetiva para objetivamente identificar e classificar a forma e composição urbana de bairros com diferentes níveis socioeconômicos, temporalidades de desenvolvimento e, conseqüentemente, diferentes composições da paisagem. O modelo criado obteve uma acurácia de 87,6%, indicando uma boa capacidade de generalização. Assim, entende-se que as características compositivas dos diferentes bairros foram efetivamente identificadas.

Limitações do estudo incluem, principalmente, elementos relacionados às imagens obtidas no GSV: (1) a visão das câmeras difere daquela de um pedestre, pois as imagens são fornecidas do centro das vias. Desta forma, elementos específicos da paisagem podem ser obstruídos, principalmente por veículos; (2) as imagens nem sempre são recentes, portanto, a possibilidade de que mudanças na paisagem possam ter ocorrido não pode ser descartada. Não obstante, a facilidade, a velocidade e os baixos custos da utilização de imagens do GSV superam consideravelmente tais limitações. Aponta-se também que, apesar da seleção de bairros com diferentes características em diversos âmbitos, o foco na classificação dos bairros em um único estudo de caso pode reduzir a possibilidade de generalização dos resultados. Estudos futuros devem buscar aplicações para a identificação de características da paisagem urbana em diferentes cenários

Apesar dessas limitações, os resultados são encorajadores como prova de conceito. A implicação prática inicial de modelos, como o proposto neste estudo, é a redução de tempo necessário para a coleta de dados. Ainda, em um âmbito teórico avança-se cientificamente na utilização de técnicas computacionais para o entendimento objetivo de qualidades urbanas, que podem influenciar o comportamento humano, bem como elementos sociais e econômicos. Estudos futuros podem explorar técnicas mais avançadas de interpretação dos resultados do modelo. Ainda, a aplicação de metodologias similares para o reconhecimento de características ambientais relacionadas a meta-qualidades urbanas como a caminhabilidade e a vitalidade.

## Notas

<sup>1</sup> <https://opencv.org/>

<sup>2</sup> <https://pytorch.org/>

## Referências

Alexander, C. (1979). *The Timeless Way of Building*, 1º ed. Oxford University Press, New York.

Amorim, R. R., Oliveira, R. C. de. (2008). As unidades de paisagem como uma categoria de análise geográfica: o exemplo do município de São Vicente-SP. *Soc. Nat.* 20, 177–198.

Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S. (2018). Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 6077–6086.

Badland, H. M., Opit, S., Witten, K., Kearns, R. A., Mavoa, S. (2010). Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits? *J. Urban Heal.* 87, 1007–1016. Disponível em: <https://doi.org/10.1007/s11524-010-9505-x>

Beidack, A. R. dos S., Fresca, T. M. (2011). Urban Restructuring and new centralities: a study about the north zone of Londrina – PR. *Bol. Geográfico* 29, 147–163. Disponível em: <https://doi.org/10.4025/bolgeogr.v29i2.9898>

Ben-joseph, E., Lee, J. S., Seoul, M., Cromley, E. K., Laden, F., Troped, P. J. (2015). Virtual and Actual: Relative Accuracy of On-Site and Web-based Instruments in Auditing the Environment for Physical Activity. *Heal. Place.* 19, 138–150. Disponível em: <https://doi.org/10.1016/j.healthplace.2012.11.001>.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, 1º ed. Springer Science + Business Media, New York, NY.

Burges, C. J. C. (2010). Geometric Methods for Feature Extraction and Dimensional Reduction - A Guided Tour, in: Maimon, Oded, Rokach, L. (Eds.), *Data mining and knowledge discovery handbook: a complete guide for practitioners and researchers*. Springer Science + Business Media, pp. 53–82. Disponível em: <https://doi.org/10.1007/978-0-387-09823-4>

Chan, D. M., Rao, R., Huang, F., Canny, J. F. (2018). t-SNE-CUDA: GPU-Accelerated t-SNE and its Applications to Modern Data. 2018 30th *Int. Symp. Comput. Archit. High Perform. Comput.* 330–338. Disponível em: <https://doi.org/10.1109/SBAC-PAD.2018.00060>

Deng, J., Dong, W., Socher, R., Li, L., Li, K., Feifei, L. (2009). ImageNet: A large-scale hierarchical image database, in: *IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL, pp. 248–255. Disponível em: <https://doi.org/10.1109/CVPR.2009.5206848>

Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A., Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A., Makes, W., Look, P. (2012). What Makes Paris Look like Paris? *ACM Trans. Graph. (SIGGRAPH 2012)* 31. Disponível em: <https://doi.org/10.1145/2185520.2185597>

Ewing, R., Handy, S. (2009). Measuring the Unmeasurable: Urban Design Qualities Related to Walkability. *J. Urban Des.* 14, 65–84. Disponível em: <https://doi.org/10.1080/13574800802451155>

Fan, H., Zheng, L., Yan, C., Yang, Y. (2018). Unsupervised Person Re-identification: Clustering and Fine-tuning. *arXiv Preprint*. Disponível em: <https://arxiv.org/abs/1705.10444>

Gehrke, S. R., Wang, L. (2020). Operationalizing the neighborhood effects of the built environment on travel behavior. *J. Transp. Geogr.* 82, 12. Disponível em: <https://doi.org/10.1016/j.jtrangeo.2019.102561>

Glorot, X., Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks, in: *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS) 2010*, Chia Laguna Resort, Sardinia, Italy., pp.249–256.

Gowda, S. N., Yuan, C. (2019). *ColorNet: Investigating the Importance of Color Spaces for Image Classification*, in: Jawahar C., Li H., Mori G., S.K. (Ed.), *Computer Vision – ACCV 2018*. Springer International Publishing, pp.581–596. Disponível em: <https://doi.org/10.1007/978-3-030-20870-7>

He, K., Sun, J. (2016). Deep Residual Learning for Image Recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp.1–9.

IBGE (2018). Panorama municipal: Londrina-Paraná [WWW Document]. URL <https://cidades.ibge.gov.br/brasil/pr/londrina/panorama> (accessed 3.26.18).

Kamalipour, H., Faizi, M., Memarian, G. (2014). Safe place by design: Urban crime in relation to spatiality and sociality. *Curr. Urban Stud.* 2, pp.152–162. Disponível em: <https://doi.org/10.4236/cus.2014.22015>

Kamilaris, A., Prenafeta-Boldú, F. X. (2018). Deep Learning in Agriculture: A Survey. *Comput.*

- Electron. Agric.* 147, pp.70–90. Disponível em: <https://doi.org/10.1016/j.compag.2018.02.016>
- Kent, J. L., Ma, L., Mulley, C. (2017). The objective and perceived built environment: What matters for happiness? *Cities Heal.* 8834, pp.1–13. Disponível em: <https://doi.org/10.1080/23748834.2017.1371456>
- Kingma, D. P., Ba, J. L. (2014). Adam: A method for stochastic optimization. *arXiv Preprint*. pp.1–15. Disponível em: <https://arxiv.org/abs/1412.6980>
- Lecun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature* 521, pp.436–444. Disponível em: <https://doi.org/10.1038/nature14539>
- Lee, S. M., Conway, T. L., Frank, L. D., Saelens, B. E., Cain, K. L., Sallis, J. F. (2017). The Relation of Perceived and Objective Environment Attributes to Neighborhood Satisfaction. *Environ. Behav.* 49, pp.136–160. Disponível em: <https://doi.org/10.1177/0013916515623823>
- Litjens, G., Kooi, T., Bejnordi, B. E., Arindra, A., Setio, A., Ciompi, F., Ghafoorian, M., Laak, J. A. W. M., Van Der Ginneken, B. Van, Sánchez, C. I. (2017). A survey on deep learning. *Medical image analysis* 42, pp.60–88. Disponível em: <https://doi.org/10.1016/j.media.2017.07.005>
- Liu, L., Silva, E. A., Wu, C., Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Comput. Environ. Urban Syst.* 65, pp.113–125. Disponível em: <https://doi.org/10.1016/j.compenvurbsys.2017.06.003>
- Maaten, L. V. D., Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9, pp.2579–2605.
- Medeiros, F. F., Grigio, A. M. (2019). Identificação das Unidades Homogêneas e Padrão da Ocupação Urbana como subsídio ao ordenamento territorial em Mossoró, RN – Brasil. *EURE (Santiago)*, 45, pp.245–270.
- Middel, A., Lukasczyk, J., Zakrzewski, S., Arnold, M., Maciejewski, R. (2019). Urban form and composition of street canyons: A human-centric big data and deep learning approach. *Landsc. Urban Plan.* 183, pp.122–132. Disponível em: <https://doi.org/10.1016/j.landurbplan.2018.12.001>
- Moosavi, V. (2017). Urban morphology meets deep learning: Exploring urban forms in one million cities, town and villages across the planet. *arXiv Preprint*. arXiv:1709, 1–10. Disponível em: <https://arxiv.org/abs/1709.02939>
- Oura, K. (2006). Verticalização em Londrina - Paraná (1950–2005): A produção do espaço urbano e seu desenvolvimento pelos edifícios verticais. *Dissertação de Mestrado - Universidade Presbiteriana São Paulo*.
- Poggio, T., Mhaskar, H., Rosasco, L., Miranda, B., Liao, Q. (2017). Why and When Can Deep – but Not Shallow – Networks Avoid the Curse of Dimensionality: a Review. *Int. J. Autom. Comput.* 14, pp.503–519. Disponível em: <https://doi.org/10.1007/s11633-017-1054-2>
- Powers, D. M. W. (2011). Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. *J. Mach. Learn. Technol.* 2, pp.37–63.
- Prefeitura do Município de Londrina (2020). Sistema de Informação Geográfica de Londrina – SIGLON [WWW Document]. Disponível em: [http://www1.londrina.pr.gov.br/index.php?option=com\\_content&view=article&id=20114&Itemid=1988](http://www1.londrina.pr.gov.br/index.php?option=com_content&view=article&id=20114&Itemid=1988) (Acessado em 15 de janeiro de 2020).
- Sallis, J. F., Cain, K. L., Conway, T. L., Gavand, K. A., Millstein, R. A., Geremia, C. M., Frank, L. D., Saelens, B. E., Glanz, K., King, A. C. (2015). Is Your Neighborhood Designed to Support Physical Activity? A Brief Streetscape Audit Tool. *Prev. Chronic Dis.* Disponível em: <https://doi.org/10.5888/pcd12.150098>
- Seresinhe, C. I., Preis, T., Mackerron, G., Moat, H. S. (2019). Happiness is Greater in More Scenic Locations. *Sci. Rep.* pp.1–11. Disponível em: <https://doi.org/10.1038/s41598-019-40854-6>
- Shen, Q., Member, S., Zeng, W., Ye, Y., Stefan, M., Schubiger, S., Burkhard, R., Qu, H. (2018). StreetVizor: Visual Exploration of Human-Scale Urban Forms Based on Street Views. *IEEE Trans. Vis. Comput. Graph.* 24, pp.1004–1013. Disponível em: <https://doi.org/10.1109/TVCG.2017.2744159>
- Shorten, C., Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *J. Big Data.* Disponível em: <https://doi.org/10.1186/s40537-019-0197-0>
- Takahashi, R., Matsubara, T., Uehara, K. (2019). Data Augmentation using Random Image Cropping and Patching for Deep CNNs. *ArXiv Preprint*. abs/1811.0, pp.1–16. Disponível em: <https://arxiv.org/abs/1811.09030>
- Tan, Y., Tang, P., Zhou, Y., Luo, W., Kang, Y., Li, G. (2017). Neurocomputing Photograph aesthetical evaluation and classification with deep convolutional neural networks. *Neurocomputing* 228, pp.165–175. Disponível em: <https://doi.org/10.1016/j.neucom.2016.08.098>

- Töws, R. L., Mendes, C. M., Vercezi, J. T. (2010). The city as a business: the case from Londrina-PR and from Maringá-PR. *Bol. Geográfico* 28, pp.91–103.
- Yamaki, H. T. (2017). *Terras do Norte: paisagem e morfologia*, 1 ed. Ed. H. Yamaki e UEL, Londrina.
- Yin, L., Cheng, Q., Wang, Z., Shao, Z. (2015). 'Big data' for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts. *Appl. Geogr.* 63, pp.337–345. Disponível em: <https://doi.org/10.1016/j.apgeog.2015.07.010>
- Yin, L., Wang, Z. (2016). Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. *Appl. Geogr.* 76, pp.147–153. Disponível em: <https://doi.org/10.1016/j.apgeog.2016.09.024>
- Yin, R. K. (2001). *Estudo de caso: Planejamento e Métodos*, 2º. ed. Bookman Companhia Editora, São Paulo.
- Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., Ratti, C. (2018). Landscape and Urban Planning Measuring human perceptions of a large-scale urban region using machine learning. *Landsc. Urban Plan.* 180, pp.148–160. Disponível em: <https://doi.org/10.1016/j.landurbplan.2018.08.020>
- Zhang, W., Li, W., Zhang, C., Hanink, D. M., Li, X., Wang, W. (2017). Parcel feature data derived from Google Street View images for urban land use classification in Brooklyn, New York City. *Data in Brief.* 12, pp.175–179. Disponível em: <https://doi.org/10.1016/j.dib.2017.04.002>
- Zhou, B., Khosla, A., Lapedriza, A., Torralba, A., Oliva, A. (2016). *Places: An Image Database for Deep Scene Understanding*. *J. Vis.* 17, pp.1–12.

## Tradução do título, resumo e palavras-chave

### *Identification of Streetscape Compositions: A Deep Learning Approach*

**Abstract.** *The environment's composition can have an impact on human behavior; however, this relationship remains uncertain until the cities' qualities and landscape can be analyzed empirically. Images obtained through Google Street View (GSV) enable a large volume of data for automated assessment of environmental characteristics. Deep learning techniques have advanced in the identification of compositional elements of the built environment. In this sense, this study seeks to investigate and test a procedure for identifying the configuration and composition of the urban landscape, classifying images obtained from GSV through a deep learning approach. From an image dataset of three different neighborhoods in Londrina-PR, a deep learning model for image classification was proposed. The model had a good performance, correctly attributing 87.6% of the samples to the corresponding neighborhoods in the case study. Compositional characteristics were empirically identified, considering the distribution of the samples in the obtained search space. The proposed model contributes to the definition of spatial units as well as in the measurement of environmental qualities, optimizing data collection, expanding sample sizes, and providing objectivity to results. This approach contributes to the expansion of city's analytical scales, identifying compositional and relational patterns in the understanding of elements influent in human behavior.*

**Keywords:** *urban morphology, built environment, deep learning, image classification, Google Street View.*

*Editor responsável pela submissão: Julio Celso Borello Vargas.  
Licenciado sob uma licença Creative Commons.*

