

## PCS - Departamento de Engenharia de Computação e Sistemas Digitais

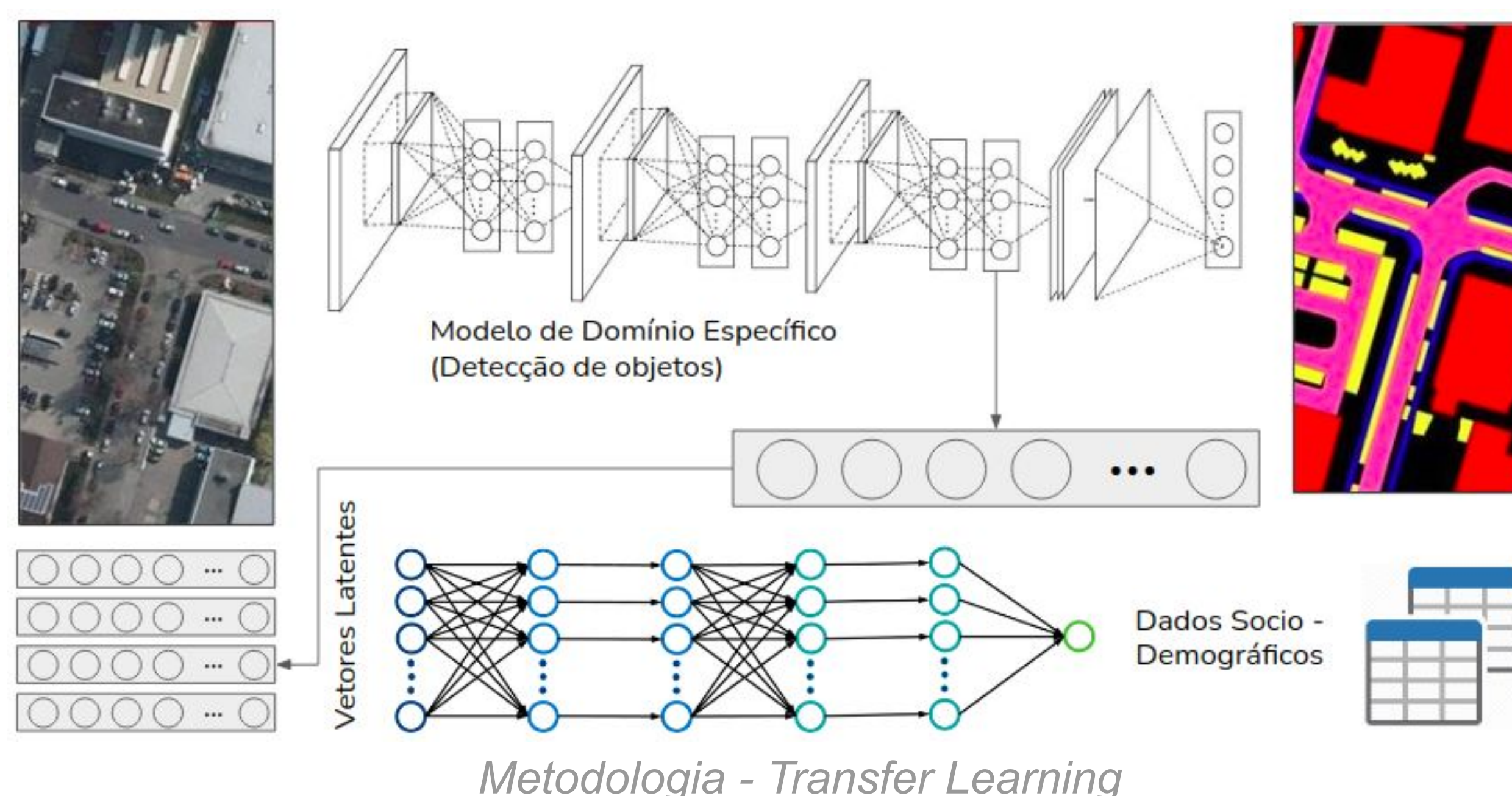
Engenharia de Computação

**Tema: Identificação de dados socioeconômicos a partir de imagens de rua e transfer learning**

### INTRODUÇÃO E MOTIVAÇÃO

O Censo demográfico feito no Brasil requer muitos recursos: cerca de 200 mil pessoas que trabalham por três meses e custa um total de R\$ 2,3 bi. Além disso, a baixa frequência da realização do censo (uma vez a cada 10 anos) torna difícil acompanhar precisamente a evolução de políticas públicas, trazendo muito incertezas quanto a resultados parciais. Com esse projeto, pretendemos apresentar uma metodologia complementar à tradicional que, apesar de dificilmente ser tão acurada, é mais barata e rápida, podendo servir perfeitamente para cálculos de valores parciais entre um censo e outro.

### METODOLOGIA



As imagens são obtidas por meio de web scraping, utilizando as informações geográficas dos censos do Reino Unido e do Brasil para buscar as imagens. Com as imagens obtidas, é feito seu processamento e o uso do modelo específico para gerar as *features* latentes de cada região e, com essas *features*, é feito o treinamento de um modelo para previsão dos deciles dos dados sociodemográficos.

### OBJETIVO

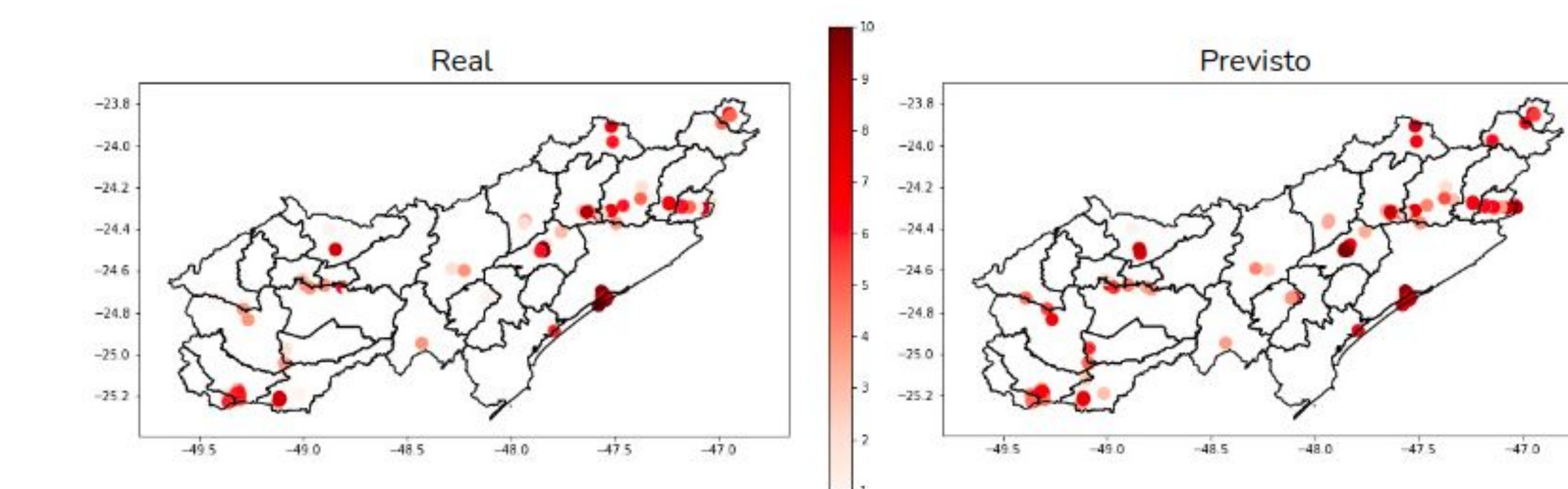


Suel, E., Polak, J.W., Bennett, J.E. et al. (2019)

### RESULTADOS E CONCLUSÃO

Apesar de algumas dificuldades na replicação, o primeiro ciclo apresentou um resultado semelhante ao original. Quanto ao segundo ciclo, embora a acurácia total foi próxima ao experimento original e sua replicação, houve uma diferença muito grande do desempenho do modelo para regiões mais ricas e urbanizadas em comparação a regiões mais pobres e ruralizadas. Em outras palavras, a metodologia utilizada apresentou uma limitação para regiões mais pobres e ruralizadas, o que não havia sido identificado anteriormente.

	Estudo	Acurácia		
		±0	±1	±2
Real	Margem de Erro	±0	±1	±2
	Londres (orig.)	0.273	0.622	0.826
	Londres (repl.)	0.234	0.541	0.734
	Vale do Ribeira	0.373	0.516	0.660



A partir do experimento criado por Suel, E., Polak, J.W., Bennett, J.E. et al. (2019), que se usa de *transfer learning* e imagens do *Google Street View* para estimar variáveis sociodemográficas em algumas regiões metropolitanas do Reino Unido, o projeto consistiu em dois ciclos: primeiramente a reprodução do experimento original e, num segundo momento, a aplicação da metodologia descrita à região do Vale do Ribeira.

**Integrantes:**

- Letícia Mancuzo de Almeida
- Willian Werner Angelo da Costa