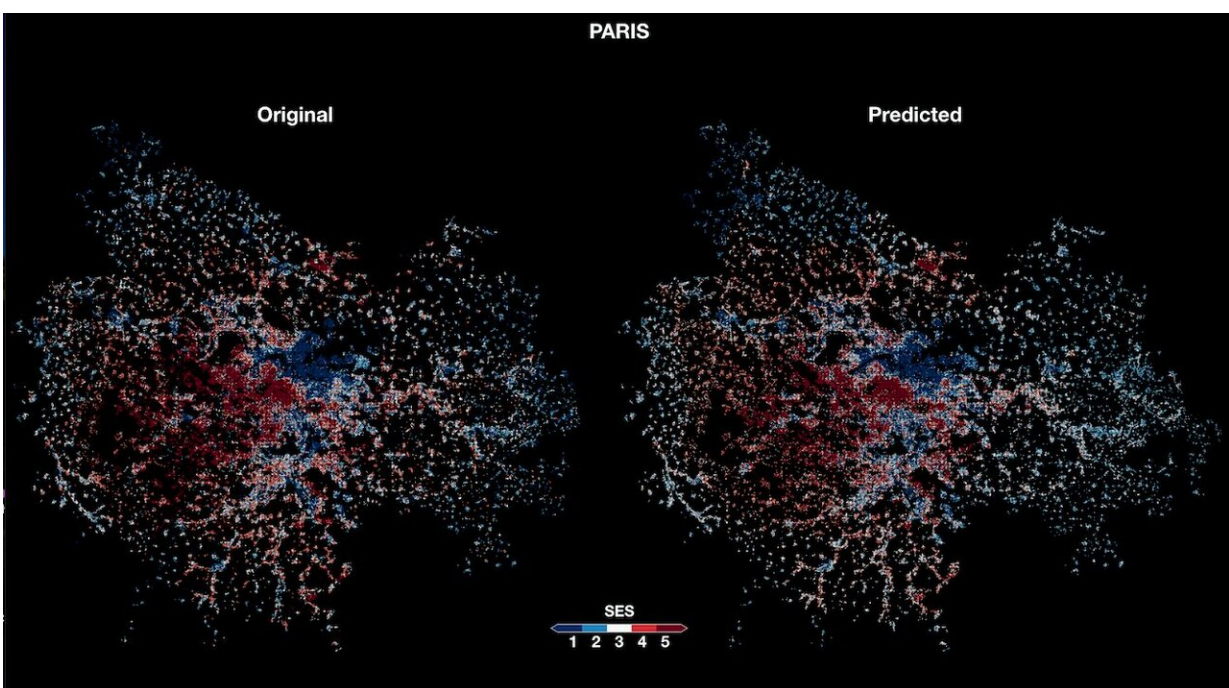


Using deep learning to infer the socioeconomic status of people in different urban areas

December 7 2020, by Ingrid Fadelli



Maps of observed and predicted average income for Paris. Each pixel represents a single 200 m × 200 m tile, with colour indicating the average SES of its inhabitants. Credit: Abitbol & Karsai.

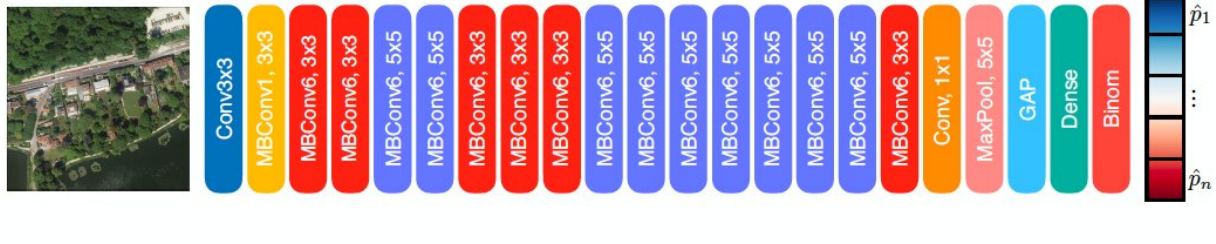
Deep learning algorithms have proved to be promising tools to tackle a variety of real-world problems, especially those that require the analysis of vast amounts of data. In contrast with other computational techniques,

in fact, these algorithms can learn to make highly accurate predictions simply by processing data related to the task they are designed to complete.

Researchers at the Ecole Normale Supérieure (ENS) de Lyon and Central European University (CEU) have recently developed a deep neural network that could be used to study the socioeconomic inequalities that can arise from urbanization. Their study, featured in *Nature Machine Intelligence*, confirms the potential of convolutional neural networks (CNNs) for the in-depth analysis of geographical regions.

For many years, efficiently tracking urbanization, the process through which an [urban area](#) becomes increasingly large and populated, has proved fairly challenging. The development of increasingly advanced remote sensing and satellite technologies, however, opened up new exciting possibilities for the observation of specific geographical regions and consequently for urbanization-related research. In their study, the researchers ENS Lyon and CEU tried to use [deep learning algorithms](#) to analyze the images collected by these tools.

"Our initial goal was actually to check what was the finest spatial resolution that we could get our algorithm (i.e., predicting the average income of an area based on its satellite image) to work with," Jacob Levy Abitbol and Marton Karsai, the researchers who carried out the study, told TechXplore. "Once we did that, we started wondering whether our underlying deep learning [model](#) was using similar features when predicting income in different cities and whether the features the model was using were ones that we would think would correlate with income or not."



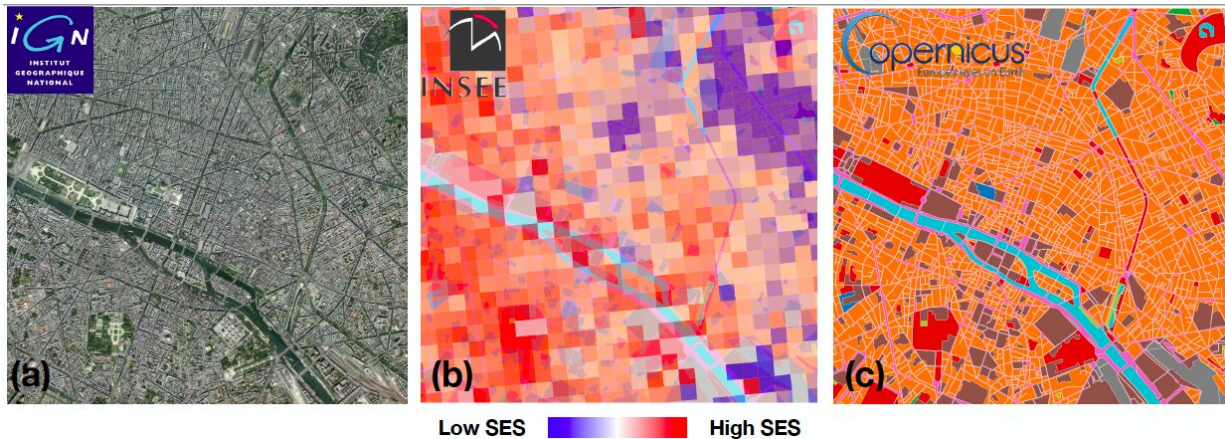
Model Architecture: The researchers' model takes the aerial tile as an input, which is then fed through several MBConv blocks. The feature maps end up going through a global average pooling layer and a dense layer to output a single value p . From it, probabilities for each socioeconomic class are generated from a binomial distribution. Credit: Abitbol & Karsai.

Abitbol and Karsai trained a CNN on aerial images of urban areas in France and evaluated its ability to predict the socioeconomic status of people inhabiting these areas. Surprisingly, when they started testing their algorithm, they found that it was activated by urban features that are not typically the most strongly correlated with income or socioeconomic status.

Over the past few years, the use of CNNs to predict the income of geographical regions based on satellite images has become fairly widespread. In order to make [accurate predictions](#), however, these models should be trained on large amounts of data, including both satellite images of the areas of interest and income-related information associated with these areas.

"The end goal of this pipeline is to use CNNs to gather new information about the economic development of a given region just by analyzing its current satellite/aerial images, without having to send people there to re-collect the census data," Abitbol and Karsai said. "It turns out that in

order for the model to do that, it would ideally need to be generalizable (i.e., if we train our model on area A it should yield consistent results on area B) and understandable (i.e., we need to know that the signals the model is using to infer that developmental data are correct)."



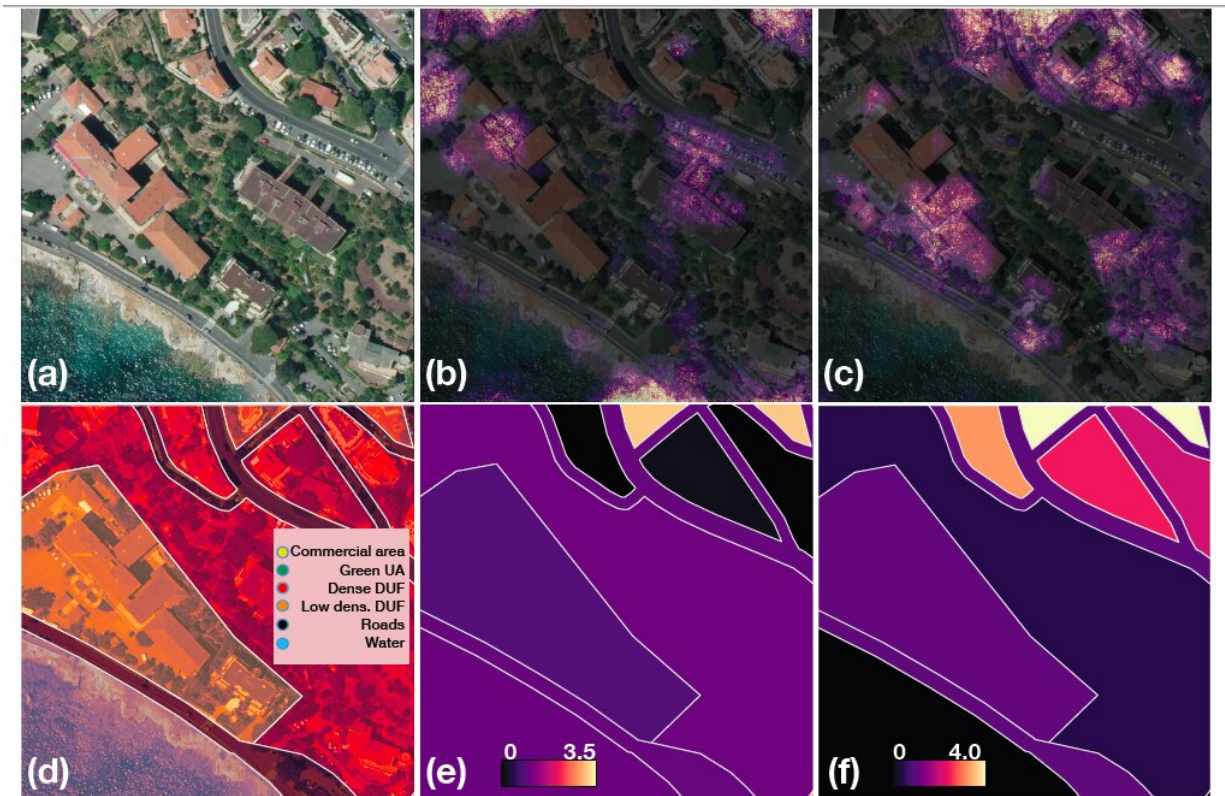
Sample of overlaid datasets (Paris). a, A 5 km × 5 km aerial tile (20 cm per pixel). b, Spatial distribution of income: each patch corresponds to a single 200 m × 200 m area with precise income data. c, Land cover map of the same area, where each color represents a different urban class. Credit: Abitbol & Karsai.

Most existing deep learning techniques that infer the average income of people in a specific area from aerial satellite images do not explain the exact processes behind their predictions. Abitbol and Karsai, on the other hand, tried to interpret their model's predictions, in order to gain a better understanding of why it inferred a specific income for each of the images it analyzed.

"When we started working on this project, some organizations were trying to launch similar models into the wild (i.e., apply them to countries where socioeconomic data is scarce to have an estimate of their development)," Abitbol and Karsai said. "The general underlying

idea was that these models were using features that were well correlated with income to perform their predictions. Our work shows that is far from the case and that we need a deeper understanding of how these CNNs assemble visual features into predictions before actually getting the most out of them."

The researchers tested their CNN on satellite data collected in different urban regions in France and found that it achieved good results. Nonetheless, they found that its predictions were based on urban features that are not generally associated with income. For example, wealthy urban areas are often characterized by the strong intensity of lights in the evening or at night, due to public spaces or commercial sites that are highly lit up, but Abitbol and Karsai found that their models primarily focused on other features, prioritizing residential areas.



Model interpretability studies using guided Grad-CAM. a–c, From an aerial tile

(a), guided Grad-CAM is used to compute activation maps for the poorest (b) and wealthiest (c) socioeconomic class. d–f, The activation maps are then overlaid with the tile's tessellation into an urban class polygon (d) to compute the normalized ratio of activations per polygon for the poorest (e) and wealthiest (f) class. UA, urban area; DUF, discontinuous urban fabric. Credit: Abitbol & Karsai.

Although CNNs have shown potential for gathering socioeconomic information about geographical regions, the study carried out by this team of researchers shows that the processes behind their predictions could be unreliable and should thus be investigated further. In the future, their work could thus inspire the development of models that can explain what features they based their predictions on, so that they can be adapted to perform more effectively and reliably.

"We would be very interested in knowing to which degree the predictions of our models and subsequent interpretation change based on the behavior of an agent that seeks to present our model with the best view of a given area, informing its predictions," Abitbol and Karsai said. "Another potential direction we are interested in involves the identification of general visual patterns, which characterize the [prediction](#) of certain income classes, to determine how much we can transfer trained models between different areas and yet gain socioeconomic predictions with high precision."

More information: Interpretable socioeconomic status inference from aerial imagery through urban patterns. *Nature Machine Intelligence*(2020). [DOI: 10.1038/s42256-020-00243-5](https://doi.org/10.1038/s42256-020-00243-5).

Citation: Using deep learning to infer the socioeconomic status of people in different urban areas (2020, December 7) retrieved 3 February 2023 from <https://techxplore.com/news/2020-12-deep-infer-socioeconomic-status-people.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.