2022

Deep learning in urban analysis for health

David Newton

Deep learning in urban analysis for health

David William Newton

College of Architecture, University of Nebraska–Lincoln,
Lincoln, NE, United States

Introduction

Over the last hundred years, the number of people living in urban areas has increased dramatically, making the link between the design of our cities and human health a pressing global issue (United Nations, Department of Economic and Social Affairs, Population Division, 2019). According to the United Nations, 56.2% of the world's population now resides in cities and the World Health Organization (WHO) projects that the number will increase to 66% by 2050. In parallel to this massive urbanization of our species, diseases such as obesity, diabetes, and high blood pressure have increased significantly among urban populations (WHO, 2016). In addition to significant physical health problems, urbanization has also been linked to increasing rates of mental illness. Analysis by the WHO estimates that mental illnesses,

Published in *Artificial Intelligence in Urban Planning and Design*, ed. Imdat As, Prithwish Basu, & Pratap Talwar (2022), pp. 121–138.
doi:10.1016/B978-0-12-823941-4.00018-4

such as depression and anxiety, make up 12% of the global disease burden, and this figure is projected to increase as more and more world's population moves into cities. The combined cost to society of treating these physical and mental diseases globally is estimated to be in the trillions of dollars (WHO, 2016). A key component in mitigating the impact of these diseases on society at large is developing a better understanding of how the design of the built environment impacts human health. Developing urban analysis methods that allow urban planners and designers to better understand this relationship is, therefore, crucial to the design of healthier cities.

The increasing urbanization of our species has also been accompanied by a revolution in information collection, processing, and analysis powered by the integration of global computer networks into the fabric of our everyday lives. From the smartphones in our hands to the global satellite networks above our heads, a vast informational infrastructure generates around 2.5 quintillion bytes of data each day. Among this deluge of data, are a multiplicity of data collection technologies that provide information on the state of the built environment. Remote sensing technologies are a major source of such data, which use aerial and satellite platforms to observe the surface of the planet from a bird's-eye view. From this vantage point, data in the form of aerial photographs, light detection and ranging (LIDAR) images, radio detection and ranging (RADAR) images, hyperspectral images, and thermal images can be captured in real-time documenting the change of the built environment over days, months, and years. These image-based datasets can be extremely large requiring new methods and technologies to systematically extract useful information. A variety of disciplines, from earth science to epidemiology, have used advances in machine learning to automate the analysis of large image-based datasets in order to better understand a range of phenomena (Tsagkatakis et al., 2019). The allied design fields, however, have largely relied on traditional inferential statistical methods—which require significant amounts of manual labor to extract features from images for tasks like regression and classification. These large image-based datasets, therefore, offer a rich and largely untapped resource for the allied design disciplines to better understand the link between human health and the built environment in a world that is becoming increasingly urbanized.

There has been significant progress in the field of machine learning in the development of methods capable of working with large image-based datasets for analysis, identification, and prediction tasks. Deep learning is a subfield of machine learning that uses layers of artificial neurons to build mathematical models from datasets that have been demonstrated to outperform competing approaches on a number of these tasks. For example, researchers have trained deep learning models to accurately identify objects, such as cars, people, exoplanets, and even skin cancer from photographs. Researchers have also begun to explore the use of remote sensing datasets, such as satellite imagery, in order to build models of natural and man-made landscapes that can aid in understanding and predicting phenomena as diverse as geological disasters and poverty (Liu and Wu, 2016; Piaggesi et al., 2019). Their application in the realm of urban analysis has been more limited, but some of the areas they have been used to analyze include: identifying urban land use (Zhang et al., 2019); predicting urban growth (Jaad and Abdelghany, 2020); and estimating human health measures, such as obesity (Maharana and Nsoesie, 2018). Their application in understanding the link between human health and the built environment has been especially limited, but existing research has demonstrated their potential for both estimating health measures, as well as identifying correlations between the visual features in the built environment and health.

The application of deep learning to urban health analysis is, therefore, in its early stages, but offers new and promising capabilities in using large image-based datasets to better understand the built environment and its effects on human health. This chapter will introduce and explore some of these capabilities, providing the allied design fields with a roadmap of this emerging area of research, its potentials, and current challenges. The chapter begins with a brief overview of existing research related to urban morphology and health, in which precedent work using traditional methods as well as deep learning are introduced. Next, research is presented demonstrating methods for the use of discriminative and generative deep learning processes for both urban health estimation and analysis. The chapter then concludes with a discussion of key challenges and directions for future work in this emerging field of research.

Urban morphology and health

Existing research in urban planning and health has established a variety of links between the physical characteristics of the built environment and human health. In terms of physical health measures, previous research has found significant correlations between characteristics of urban morphology, such as density and street network pattern, to rates of obesity (Lopez- Zetina et al., 2006; Marshall et al., 2014). These characteristics have also been found to be linked to increased rates of diabetes (Marshall et al., 2014) and asthma (McConnell et al., 2006). The work done so far suggests that neighborhoods and cities that are more walkable tend to be correlated with improved health outcomes for the diseases mentioned.

There has also been a growing body of work that has discovered significant correlations between urban morphology and mental health. Research looking at how urban density effects mental health has found a positive correlation between high rates of urbanization and high rates of mental illness (Peen et al., 2010). Street network proximity has been linked to neurological diseases such as non-Alzheimer's dementia, Parkinson's disease, Alzheimer's disease, and multiple sclerosis (Yuchi et al., 2020). In contrast, proxies for lower density, such as access to green spaces, water features, natural views, and natural light, have been found to correlate to low rates of anxiety and depression (Braubach, 2007; May et al., 2009; Garrett et al., 2019).

The picture that emerges from this growing body of work is that the physical characteristics of our neighborhoods and cities have significant correlations with health. The nature of these correlations is still being studied, and it is important not to confuse correlation with causation, but the evidence suggests an important link that requires more investigation. The majority of existing research in this area has primarily used traditional inferential statistical approaches to discover correlations (Hoisington et al., 2019; Renalds et al., 2010). These approaches, however, have a limited ability to efficiently analyze large image-based datasets, such as those from remote sensing platforms.

Deep learning in urban analysis for health

In order to address some of the shortcomings of traditional statistical methods, there is a growing body of research investigating the use of deep learning in combination with remote sensing datasets to discover and better understand correlations between urban morphology and human health. This research can be broadly categorized into two categories: discriminative deep learning and generative deep learning approaches. Deep learning models are comprised of layers of artificial neurons—with each neuron being a simple mathematical function mapping inputs to an output. These simple building blocks can be connected to one another in networks in order to create models capable of representing any mathematical function. The organization of multiple layers of artificial neurons into a network to accomplish a particular task is referred to as creating a deep learning architecture. There are a large variety of architectures for both discriminative and generative deep learning that has been developed and validated by the research community with new architectures being developed every day.

Discriminative deep learning processes use labeled datasets to build models for classification and regression tasks. In cases where an input dataset is correlated with an output dataset, these processes can approximate a function that maps inputs to outputs given enough data examples and training time. Generative deep learning processes work in a different way and use large unlabeled datasets to learn the probability distribution that underlies an input dataset. This distribution can then be sampled to generate new data instances. Generative processes require less data preparation than discriminative processes because they work with unlabeled data. Discriminative and generative processes can build models from many types of large datasets (e.g., images, drawings, text, 3D models, sounds, etc.). This flexibility, coupled with their ability to work with images, makes them useful to disciplines whose data tends to be image-based, or heterogeneous, in nature.

Applications of discriminative deep learning in urban health analysis

Discriminative deep learning approaches have been most widely used by existing research in urban health analysis. They have been used in conjunction with aerial, satellite, and point-of-view images for a variety of classification and regression tasks involving demographics, health, and well-being. For example, they have been used to train models that can estimate the population of census blocks from satellite images using classification (Robinson et al., 2017). They have also been applied to regression tasks to estimate the rate of poverty in developing countries using daytime and nighttime satellite images of those countries (Jean et al., 2016). In terms of health measures, researchers have trained discriminative models on satellite images of cities to estimate rates of obesity (Maharana and Nsoesie, 2018). Researchers have also used street view and point-of-view images to estimate a broader spectrum of wellness metrics related to unemployment, education, income, and wellbeing (Suel et al., 2019).

Convolutional neural networks (CNNs) are a deep learning architecture developed for working with images and are the main architecture used by this precedent research. CNNs work by taking image data as an input and passing that data through a series of neural layers. As the images move through each layer, image data is progressively abstracted into sets of visual features that provide a compressed representation of the image data that can be used for classification, regression, or generative tasks. The layers at the beginning of the model extract low-level features (i.e., edges, corners, etc.) while the layers toward the end of the model extract high-level features (i.e., roads, buildings, etc.). CNNs learn which features best define an image for a particular task and how to extract those features from the image data through a training process involving feeding example images into the model along with the desired model output (e.g., a desired classification or regression value), calculating the error, and then using an optimization algorithm to adjust the weights associated with the CNN's mathematical model. This supervised learning process is done iteratively until the model reaches peak accuracy.

There are a large variety of CNN architectures to choose from depending on the task at hand. Previous research involving the use of satellite images for health analysis has primarily used the visual geometry group (VGG) family of CNN architectures (Simonyan and Zisserman, 2016). There are, however, a number of other architectures that offer increased accuracy in image recognition tasks that could also be chosen (e.g., Inception, Xception ResNet). Training these CNN models, however, poses a challenge. Discriminative deep learning models require prodigious amounts of data for their training. These models are often trained on datasets that contain millions of data samples in order to reach peak accuracy in classification, or regression tasks. This can pose a challenge when working with smaller datasets.

In order to address this challenge, researchers have developed two methods that are fundamental to any deep learning training process: data augmentation and transfer learning. Data augmentation increases the size of the training dataset by creating new data instances from existing instances. In the case of an image dataset, this is done by taking an existing image from the dataset and applying operations (e.g., scaling, rotation, distorting, adding noise, etc.) to it that modify the image from its original state. The new modified image can then be used as a new training example. This simple trick seems dubious but has been demonstrated to improve model accuracy significantly and is used extensively by precedent research in urban health analysis.

Transfer learning is the other primary method used when working with small datasets (i.e., datasets in the hundreds to thousands of data points). Transfer learning saves significant computation time by repurposing deep learning models trained for one task for another similar task. This is done by using available deep learning models trained on millions of data points and then retraining only a small part of that model for the desired classification, or regression task that is similar but different from the task the model was originally trained for. Transfer learning has demonstrated impressive capabilities and allows the analytic insights developed from one dataset to be transferred to other datasets. Precedent work in urban health analysis has made use of this method extensively.

Figure 1 shows an example of discriminative deep learning architecture using transfer learning for an urban health regression task involving estimating the rate of overweight adults based on satellite images of US census tracts—which are typically about the scale of a neighborhood. In the figure, the Xception CNN architecture is

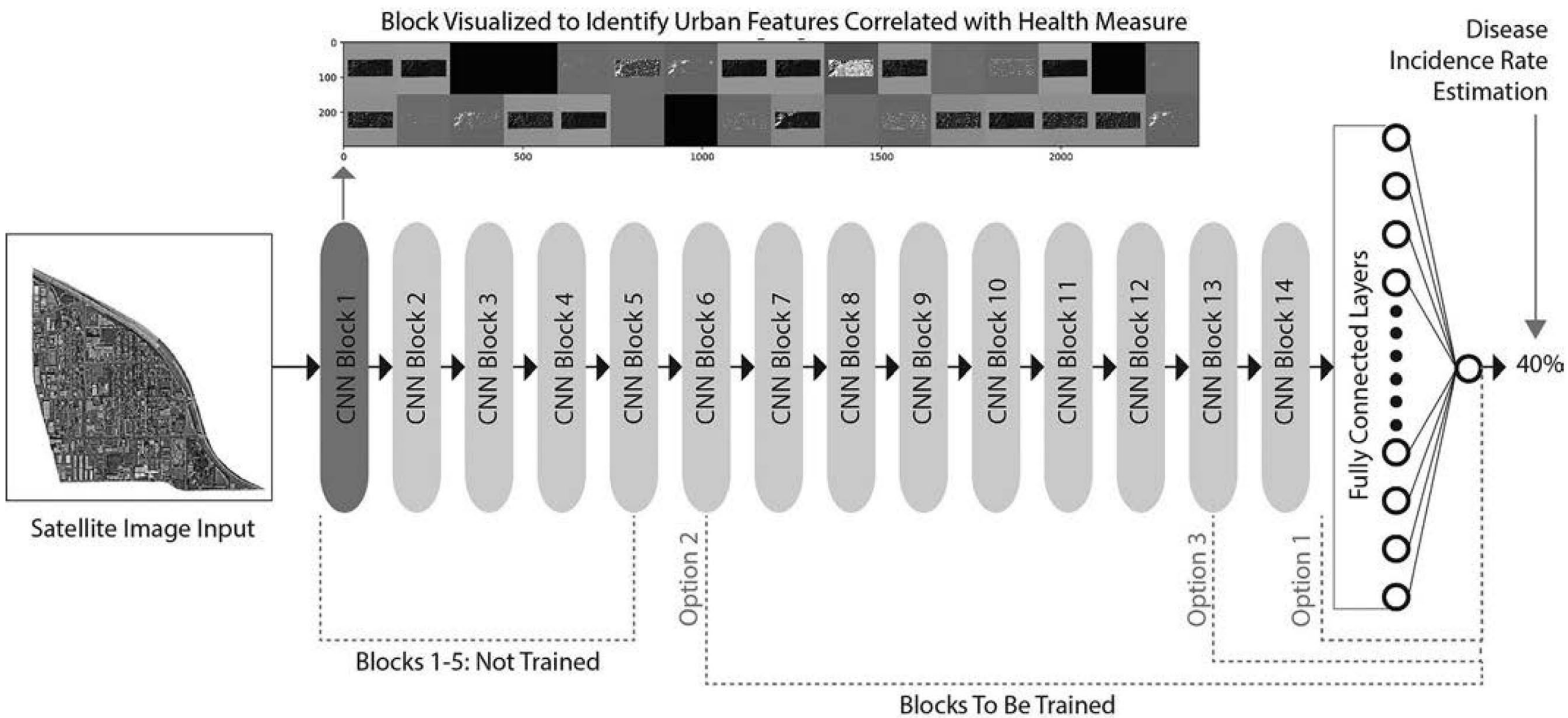

Figure 1 Xception CNN architecture. The Xception CNN architecture is shown. The diagram shows that the following CNN blocks are trained on the dataset: Option 1—just the last fully connected layer; Option 2—blocks 13, 14, and the last layer; Option 3—CNN block 6–14 and the last layer. The diagram shows that block 1 is visualized to find correlations between specific urban design features and the estimated health outcome.

pictured. Xception is an architecture that is pretrained on the ImageNet database, which is an image database of over 14 million images spanning more than 20,000 object categories. The original architecture is comprised of 14 convolutional blocks (each block is made up of several neural layers) and a layer at the end that outputs a classification value. In order to adapt the model for estimating rates of obesity, the final layer of the original model is removed and replaced by a new layer that will output a regression value instead of a classification value. As Figure 1 shows, the model takes satellite images of census tracts as an input, extracts features through its convolutional blocks, and then outputs a regression value estimating the rate of overweight adults in the census tract.

The training of the model involves freezing a set number of neural layers and only training a select set of layers in the model. This selective training is what saves time and finding which layers to train is a key problem. This choice is often made based on how similar the desired dataset is to the original dataset for which the model was trained. In the example given in Figure 1, satellite images are significantly different than the ImageNet database used to train the original Xception architecture. ImageNet features close-up elevational views of various objects (e.g., people, plants, animals, furniture, etc.) and not views from above. In order to address this issue, multiple options are normally tested. For example, one test might explore how well a minimally modified version of the Xception architecture can perform by only training the last layer. This option saves the most computation time but assumes that the low and high-level image features learned from ImageNet will be relevant for analyzing satellite images of cities. The second test might explore the hypothesis that the low-level image features learned from ImageNet are useful but that the high-level features are not relevant. Therefore, this approach might train the last two convolutional blocks of the Xception architecture as well as the last layer. The third test might explore the hypothesis that both high-level and a proportion of low-level features may not be relevant for satellite image analysis for a particular health measure. Therefore, convolutional blocks 6 through 14, as well as the final layer, may be trained. As more layers are trained, more computational resources and time for that training are necessary. In the example, because aerial views are significantly different than elevational views, the third architecture demonstrated the lowest error in estimating rates of overweight adults but required the largest computational resources.

The existing research presented in this section establishes the efficacy of discriminative methods for estimating some health and well-being measures but there are still a number of areas that require additional study in order to realize the full potential of these processes for urban health analysis. These areas include the following: developing a greater understanding of which health measures can best be estimated with these processes; creating methods for training these models more efficiently: and developing techniques to identify specific visual features that correlate with health measures. The next section will address this last issue in more detail.

Analyzing deep learning models to find correlations

Deep learning models are often referred to as "black-box" models because their inner workings remain obscured behind hundreds of thousands, and sometimes millions, of parameters. The development of analytic methods to address this problem is currently a pressing problem for disciplines working with deep learning because such methods would allow insight into the learned correlations between dataset features and estimation values. Previous research in this area has used the visualization of individual CNN layers to identify correlated features. This approach has been used extensively in work using satellite images to estimate health (Jean et al., 2016; Maharana and Nsoesie, 2018) but has significant drawbacks. Specifically, these methods rely heavily on visual interpretation to identify features of interest and provide little information on how combinations of features might be correlated with outcomes.

Researchers in the field of machine learning have developed a variety of methods to identify possible correlations between dataset features and predicted outcomes in deep learning models. Zeiler and Fergus (2014) have developed a quantitative method involving deconvolution that highlights the portion of an image that is being activated by a particular neural unit. Nguyen et al. (2019) have used optimization techniques to find images that cause the highest and lowest activation of different neural layers. Gatys et al. (2016) have used the calculation of Gram matrices to find the neural layers most activated by a given set of images. The identified layers can then be visualized as images called feature maps that can be interpreted by an analyst to identify key visual features.

Figure 2 shows an example of this last approach. In the example, a dataset of satellite images of census tracts from the state of California is first subdivided into image sets that represent high and low incidence for three different health measures: obesity, asthma, and heart disease (Newton, 2021). The average Gram matrix is then calculated for each high and low incidence image set. This is done by calculating the Gram matrix for each individual census tract image from the first convolutional block of the Xception architecture for each set and then averaging those individual Gram matrices together. Figure 2 shows visualizations of the average Gram matrix calculated for each health measure. These matrices serve as a kind of spectrograph for the satellite images present in each high and low incidence set and allow each health measure to be compared. For example, obesity and heart disease show a similar pattern of activation for high incidence images, while asthma is noticeably different.

The axes of the matrix show identification numbers for the specific neural layers (i.e., feature maps) in the first convolutional block. Bright colors in the Gram matrix represent combinations of feature maps that are most active on average for a particularly high or low incidence set. These feature maps can then be visualized and interpreted to identify specific built and natural environment features that are correlated with high and low incidence rates. Gram matrices can be calculated from any convolutional block in the CNN architecture, and the choice of where to do this is an important one. For this example, the first convolutional block was chosen because it allowed for easier visual interpretation. The downside to this choice is that the neural layers at this level are involved with identifying low-level image features (e.g., edges, corners, etc.) and not high-level features (e.g., objects composed of several low-level features like street network grids, etc.).

In **Figure 3**, the most active feature map combinations identified from the average Gram matrices for both high and low incidence rates of overweight adults are shown (Newton, 2021). The highest

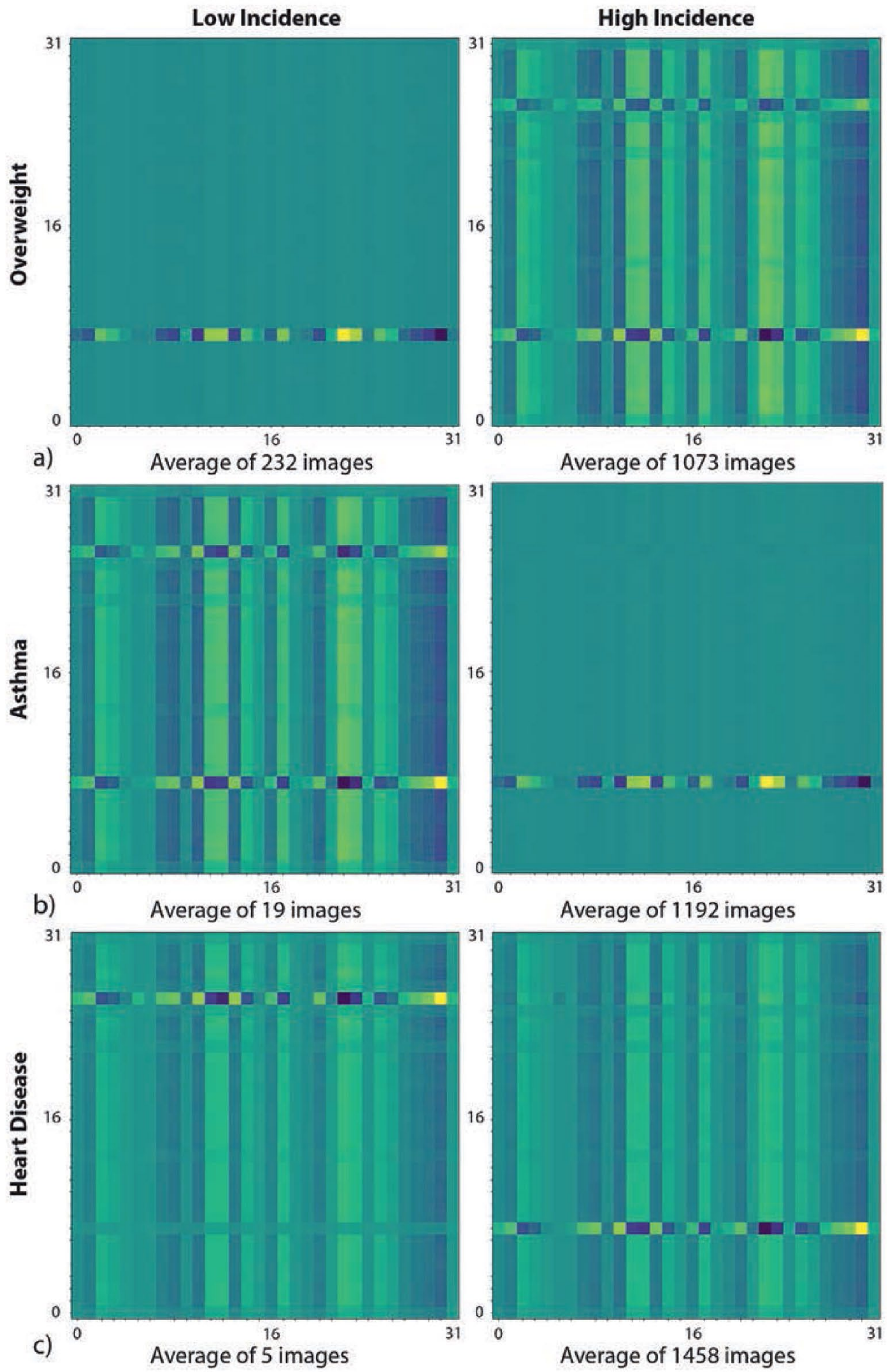

Figure 2 Gram matrices. (A) Shows average Gram matrices of low and high incidence census tracts for overweight health measures. (B) Shows average Gram matrices for asthma. (C) Shows average Gram matrices for heart disease.

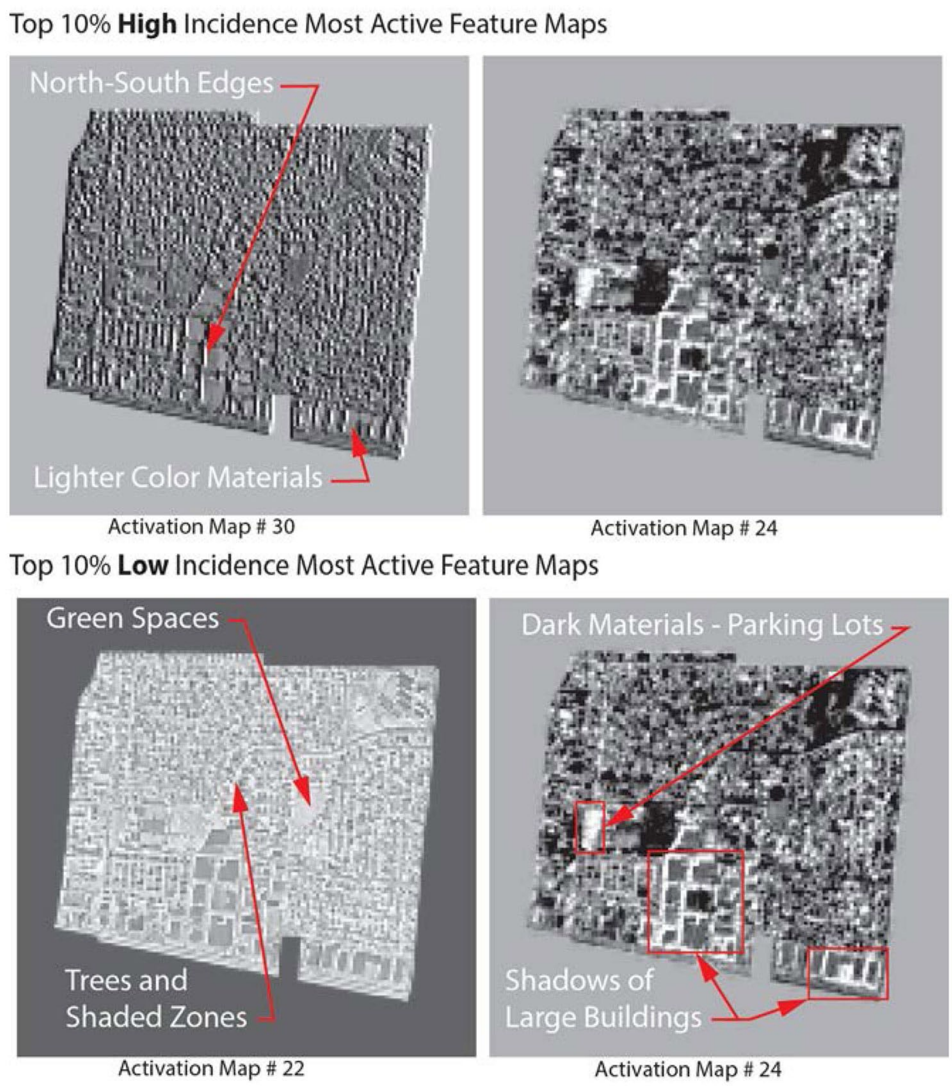

Figure 3 Gram matrix analysis. A sample census tract image with a high rate of overweight adults is used as the input to the CNN model. The most active feature maps identified by the Gram matrix analysis are shown for both high and low disease incidences. Qualitative analysis is overlaid on the feature maps to identify specific visual features that are activating the model. Brighter pixel values indicate more activation in that area of the image.

activating feature maps for high disease incidence are feature maps 30 and 24. Figure 3 shows visualizations of these feature maps as well as overlaid analysis. Visual analysis of feature map 30 reveals that it activates most when detecting proxies for buildings and streets—such as north-south edges and the roofs of buildings—especially lighter roofing materials often associated with larger commercial and residential buildings. In contrast, feature map 24 activates in relation to the space in-between buildings, specifically darker elements in the exterior landscape of the census tract, such as asphalt surfaces (e.g., streets and parking lots), vegetation, and shadows. The highest active feature maps for low incidence are feature maps 24 and 22 as shown in Figure 3. Feature map 24 is the most active for both high and low incidence rates. Feature map 22 has a similar activation behavior as 24, responding to exterior spaces. This activation pattern, therefore, focuses more on exterior space than what was seen in the high incidence case. These results indicate the CNN model is most responsive toward proxies for walkability, such as streets, shadow patterns along streets, and parking lots. These results are consistent with precedent research in urban planning and health that has found similar correlations between walkability and obesity, but this method provides a new way of identifying these correlations (Li et al., 2009; Marshall et al., 2014).

This example demonstrates a deep learning-driven mixed methods approach to identify correlations between satellite image features and disease incidence and also its limitations. The first major limitation involves the selection of where in the CNN model to calculate the Gram matrices and retrieve the feature maps. In this research, the first convolutional block was chosen because, at that stage in the model, images can still be readily interpreted through visual examination. The first blocks of the CNN model have neural layers that learn to find low-level features. While using these early layers from the analysis allows the feature maps produced by these layers to be human-readable, the feature maps at this stage have learned only very basic representations. This makes developing insight about how high-level features (e.g., street grid patterns, park distribution patterns, building density differences, etc.) are correlating with specific outcomes more difficult and subject to a greater level of interpretation. This issue relates to another limitation, which is the degree of interpretation needed to interpret the activation patterns at work in the feature maps identified by the Gram matrices. Identifying the image features that are activating a particular feature map requires a careful assessment of the feature maps on a pixel-by-pixel basis. For some feature maps, the activations can be straightforward to interpret, but others require a greater degree of subjective judgment. Developing more robust quantitative methods for the analysis of CNN models to identify these features is, therefore, a pressing issue that has become the focus of an area of research called explainable, or interpretable, artificial intelligence. Recent work in this area has shown significant improvements over previous work (Linardatos et al., 2021), and with new developments occurring each year, robust tools to address this issue seem within reach.

Applications of generative deep learning for urban health analysis

Generative deep learning processes have been explored less than discriminative processes for urban health analysis by existing research. This may be due to the fact that they do not offer straightforward classification, or regression, values that lend themselves to quick interpretation, but instead can learn the statistical correlations that define one type of dataset versus another and can create new data instances based on these learned correlations. There are a wide variety of generative deep learning architectures that have been developed in the field of machine learning, such as variational autoencoders and deep belief networks. Generative adversarial networks (GANs) proposed by Goodfellow et al. (2014) are the most popular deep generative model. This popularity is due to their ability to outperform competing approaches in terms of their flexibility for image generation tasks and the quality of produced images.

GANs work through the competition of two deep neural networks: the generator network and the discriminator network. The job of the generator network is to create new data instances from noise. The job of the discriminator network is to correctly identify the fake images being created by the generator network from the real images comprising the training dataset. Both networks are trained together in an iterative manner and, if the training process is successful, the generator will gradually learn to produce new data instances that are good enough to fool the discriminator network. **Figure 4** shows an example of GAN architecture illustrating this deep learning architecture. In the figure, the GAN is being trained on a dataset of satellite images of

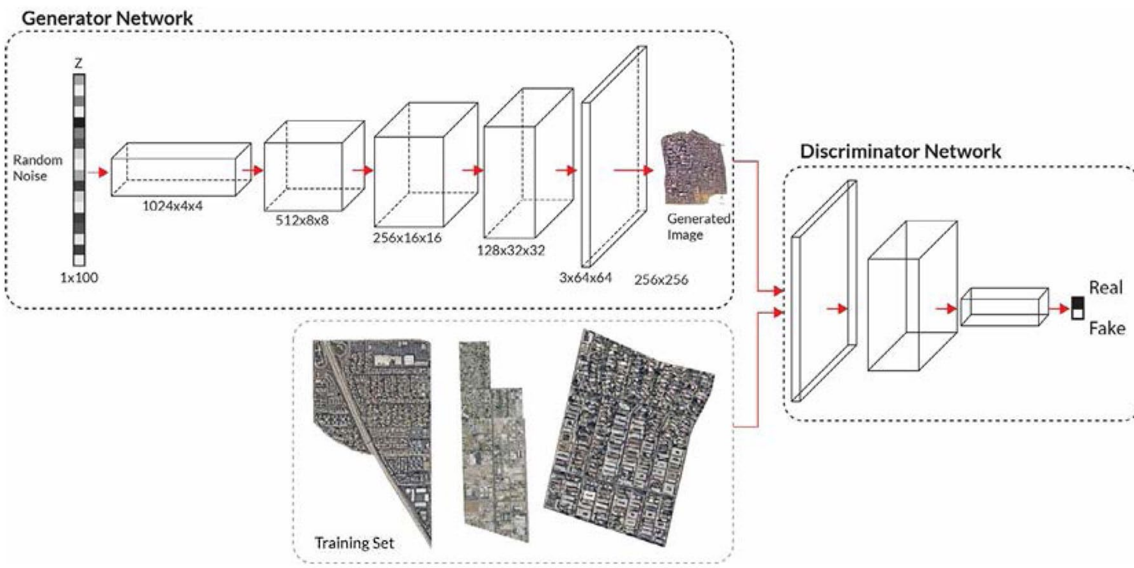

Figure 4 GAN architecture. The GAN architecture is comprised of a generator and discriminator network that compete against each other. Through this competition, the generator learns the probability distribution that underlies a training set of images and can learn to create new image instances by sampling that distribution.

census tracts. The generator network is tasked with learning to generate completely new images that resemble those in the training set. The discriminator network must, therefore, learn to accurately differentiate between data instances from the real training set and those being artificially created by the generator.

GANs have been used for a large variety of image generation tasks. For example, they have been used to generate images of human faces, bedroom layouts, and building facades. They have also been used in the creation of designs for new 3D objects like chairs and tables. Their application for urban health analysis, however, has been limited. The research that has been done in this area can be classified as residing in two different categories: (1) approaches that use GAN architectures to create completely new data instances for analysis; (2) approaches that use GAN architectures for translation between one dataset and another for analysis.

The first category involves training GANs on satellite, aerial, or map images of exemplar urban areas in order to create images of new urban plans that have been learned from the exemplar dataset. These images can then be qualitatively assessed to develop insight into

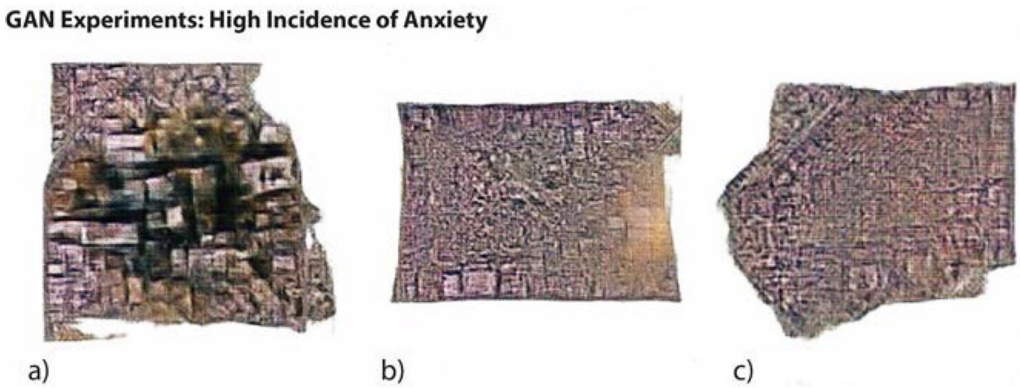

Figure 5 High anxiety GAN model. Shows generated samples from a GAN model trained on the high incidence of anxiety. Parts (A)–(C) show dense urban fabric with no natural spaces and an air pollution-like haze.

correlations that might underpin the given set of exemplar designs. An example of this kind of approach is shown in **Figure 5**, where a GAN is trained on satellite images of census tracts with high rates of anxiety in order to generate new census tract designs that may be correlated with that health measure (Newton, 2020). A qualitative visual analysis of these generated images reveals urban design features that may be correlated with high rates of anxiety. In part (A) of the figure, large, shaded areas can be seen indicating limited accessibility to natural light. In part (B), a dense urban grid is seen within an orange pollution-like haze. In part (C), a diagonal line looking like an airport runway or highway interrupts the density of the urban grid. The generated images in parts (A)–(C) of Figure 5 are notable in the high levels of density they show with no green spaces or natural landscapes. In contrast, **Figure 6** shows the results of training a GAN on census tract images with low levels of anxiety. These images indicate that natural landscape features (e.g., green spaces, open land, mountains, water features), access to light, and medium to low density may have a meaningful correlation with low anxiety rates.

The other category of approach involves the use of GANs to train models that can translate from one set of exemplar images to another set. These translations can then be studied to identify urban design features that might be correlated with a particular outcome, such as safety, or health. For example, researchers have used this method to translate between satellite images of areas with a high incidence of

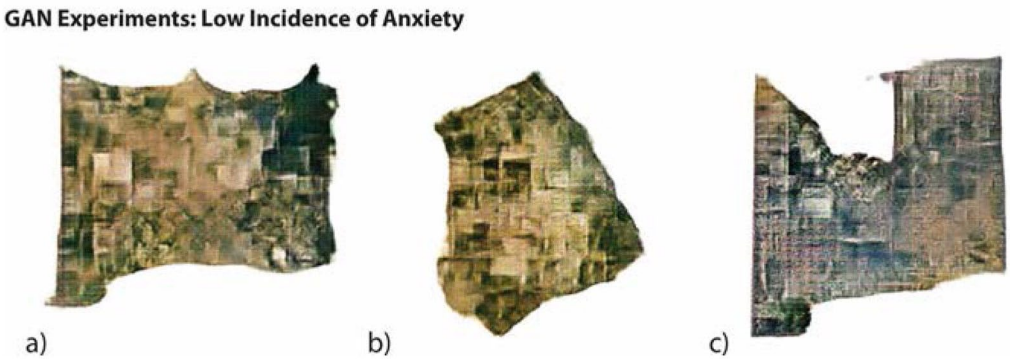

Figure 6 Low anxiety GAN model. Shows generated samples from a GAN model trained on the low incidence of anxiety. Parts (A)–(C) show census tracts dominated by natural landscapes.

bicycle accidents to areas of low incidence in order to identify urban features (e.g., street design, sidewalk design, etc.) that may be correlated with lower accident rates (Zhao et al., 2019).

Figure 7 shows an example using the CycleGAN architecture for an analysis task involving depression. Parts (A), (D), and (G) of the figure show an original satellite image of a California census tract with a high incidence of depression. Parts (B), (E), and (H) show a translation of the original satellite image by the CycleGAN to be more consistent with the image feature present in low incidence images. Parts (C), (F), and (I) show a pixel-by-pixel difference between the original and the translated image—highlighting the primary features that have been changed. A qualitative visual analysis of these GAN translation results shows changes to a street grid pattern and greenspace distribution.

Comparing the results of both GAN studies to existing research can help to validate the potential correlations identified. In terms of anxiety, existing research has found that increased levels of urbanization and pollution correlated with higher rates of mental illness (Bolton et al., 2013; Chen et al., 2018; Peen et al., 2010). Further, correlations between anxiety and exposure to natural light and natural views have also been identified (Braubach, 2007; May et al., 2009). In relation to depression, previous research has also shown a significant correlation between low incidences of depression and access to greenspaces (Beyer et al., 2014; May et al., 2009; Cohen-Cline et al., 2015;

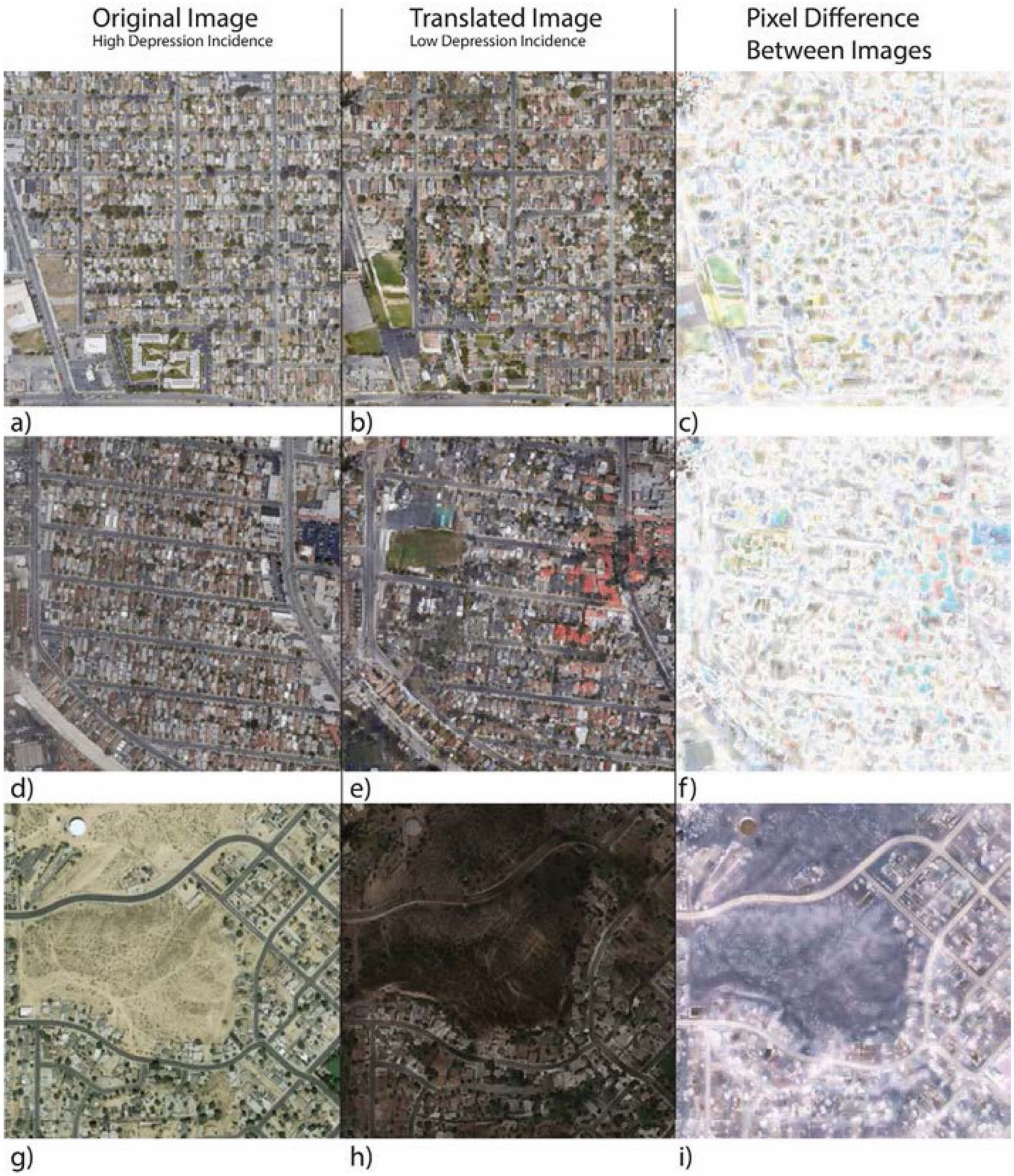

Figure 7 Depression GAN model. (A, D, G) shows an original satellite image of a California census tract with a high incidence of depression. (B, E, H) shows a translation of the original satellite image by the CycleGAN to be more consistent with the image features present in low incidence images. (C, F, I) shows a pixel-by-pixel difference between the original and the translated image—highlighting the primary features that have been changed.

Rautio et al., 2018). The results of these GAN experiments, therefore, are consistent with findings from existing research, but limitations inherent in this mixed-methods process need to be addressed to better verify these results. These limitations mostly stem from the qualitative visual analysis used to identify correlations. This process involves a significant degree of subjective interpretation and also does not provide detailed information on the nature and degree of the correlation between identified features and health outcomes. Integrating additional quantitative statistical methods (e.g., Pearson Correlation, etc.) to validate identified correlations is one possible way of addressing this issue. Other approaches involve developing quantitative methods of analyzing GAN models that can identify which learned visual features generated by a GAN architecture correlate most with a specific health outcome.

Existing research in this area is investigating how generative models might be used for urban health analysis, but as with discriminative deep learning processes, there are still many open questions involving how these models might be used to identify specific design features correlated with health measures, as well as, developing methods of building datasets and training models that are the most efficient.

Challenges, opportunities, and next steps

The existing research and examples presented demonstrate the potential efficacy of using deep learning with remote sensing data for urban health analysis tasks, but there are a number of important challenges that will need to be addressed by future research in order to realize the full potential of this technology to illuminate the links between the built environment and human health for the design disciplines. These challenges reside in four key areas which will be discussed in more detail below: (1) overcoming the high entry barrier to using deep learning; (2) acquiring and prepping the necessary data for deep learning; (3) developing efficient methods to train deep learning models for urban analysis; and (4) moving from an understanding of correlation to one of causation.

A key challenge in working with deep learning models is overcoming the high entry barrier needed to effectively train and analyze these models. This challenge is especially acute in the design disciplines, where knowledge of programming and machine learning is rare. Developing competencies in these areas is therefore key for the allied design disciplines if they are to more fully engage the potentials of current and future machine learning technologies for urban analysis. There are a large variety of massive open online courses in addition to publications introducing machine learning that can work as an effective stop-gap to build basic competencies in these areas, but a more strategic approach would better position the allied design fields to shape the future development and use of these technologies in the analysis of the built environment. One example of a more strategic approach could be integrating core competencies in programming, data science, and machine learning with design curriculums. This would give future practitioners and researchers in the allied design fields the foundation they need to effectively lead discussions and develop methods to extract critical insight on human health and other factors from the ever-expanding streams of data produced on the built environment.

Another important challenge in working with deep learning for urban health analysis involves acquiring the necessary remote sensing and health data to train deep learning models. Accessing quality health data for deep learning is a challenge due to the cost of collecting accurate health data and the security measures that are necessary for protecting individual privacy. Existing deep learning research involving health data has primarily used bulk anonymized data from governmental sources. These datasets are often limited in terms of geographic coverage and scale. In the United States, for example, health data is usually recorded at the county level, while data at smaller scales (e.g., census tract, neighborhood scale, etc.) is often not available. In terms of remote sensing datasets, the majority of existing research uses satellite, or aerial, images. These images, however, comprise only one data stream among many other remote sensing datasets (e.g., LIDAR images, RADAR images, hyperspectral images, thermal images, etc.) that could be useful for urban health analysis. There are also nontraditional remote sensing datasets, such as the use of social media streams that have demonstrated efficacy for urban analysis (Frias-Martinez and Frias-Martinez, 2014). These datasets are widely accessible through private (e.g., Google Earth, Bing Maps

Imagery, etc.) and public sources (SGS Earth Explorer, NASA Earthdata Search, DigitalGlobe Open Data Program, etc.). The principal challenge in working with this data, therefore, is deciding which data streams might be most efficacious for a particular health analysis task and also in preparing the data (e.g., removing incomplete/damaged data instances, cropping input images consistently, etc.). The field of machine learning has attempted to address this problem in relation to classification problems through the creation of standardized datasets (e.g., ImageNet, MINST, ModelNet, etc.) that are easily accessible to the research community. These shared datasets allow researchers to save time in data collection and preparation, while also providing a more robust ability to directly compare the results of one research project to another. Developing standardized datasets for urban health analysis is, therefore, crucial for future research in this area.

The next challenge is that the training of deep learning models can be very resource-intensive—requiring large amounts of data and computing time. As discussed previously, transfer learning can dramatically reduce the amount of data and computing resources needed to train a deep learning model by making use of pretrained models trained on other image datasets that are similar in scale and view angle to a target image dataset (e.g., aerial images, LIDAR images, etc.). The problem is that remote sensing images are often very dissimilar in scale and view angle when compared to the images used to train available pretrained deep learning models. This dissimilarity makes it less efficient for transfer learning. In order to address this issue, a library of pretrained deep learning architectures is needed that are trained on remote sensing datasets. These pretrained models should be provided for different remote sensing data types, such as satellite, thermal, and hyperspectral images.

The last key challenge involves moving from an understanding of the correlation between health and the built environment to an understanding of causation. Establishing causation means demonstrating that a particular health outcome came as a consequence of some design feature in the built environment and not by chance, or due to some other hidden factor. This kind of work requires significant monetary investment and, therefore, a renewed sense of urgency by governments to prioritize research on the built environment is needed. In order to foster this kind of attention, a compelling evidence-based case must first be made correlating the built environment with human health. Deep learning workflows could provide the means to help build this case.

Addressing these challenges could allow for a new era of public health analysis and land use planning. One in which the capabilities of deep learning are used to better understand and predict the relationship between human health and its physical environs from a variety of data streams generated in our neighborhoods and cities. These predictive models could provide significant cost savings for countries around the world and help them to better deal with emerging health crises, such as pandemics. The stakes are, therefore, very high, and it is more pressing than ever that the challenges outlined be addressed in order to realize a new data-driven era of planning for our built environments.

References

Beyer, K., et al., 2014. Exposure to neighborhood green space and mental health: evidence from the survey of the health of Wisconsin. Int. J. Environ. Res. Public Health 11 (3), 3453–3472. https://doi.org/10.3390/ijerph110303453 MDPI AG.

Bolton, J.L., et al., 2013. Maternal stress and effects of prenatal air pollution on offspring mental health outcomes in mice. Environ. Health Perspect. 121 (9), 1075–1082. https://doi.org/10.1289/ehp.1306560 Environmental Health Perspectives.

Braubach, M., 2007. Residential conditions and their impact on residential environment satisfaction and health: results of the WHO large analysis and review of European housing and health status (LARES) study. Int. J. Environ. Pollut. https://doi.org/10.1504/IJEP.2007.014817 Germany: Inderscience Publishers.

Chen, C., et al., 2018. Ambient air pollution and daily hospital admissions for mental disorders in Shanghai, China. Sci. Total Environ. 613–614, 324–330. https://doi.org/10.1016/j.scitotenv.2017.09.098 Elsevier BV.

Cohen-Cline, H., Turkheimer, E., Duncan, G.E., 2015. Access to green space, physical activity and mental health: a twin study. J. Epidemiol. Community Health 69 (6), 523–529. https://doi.org/10.1136/jech-2014-204667 BMJ.

Frias-Martinez, V., Frias-Martinez, E., 2014. Spectral clustering for sensing urban land use using Twitter activity. Eng. Appl. Artif. Intell. 35, 237–245. https://doi.org/10.1016/j.engappai.2014.06.019 Elsevier BV.

Garrett, J.K., et al., 2019. Coastal proximity and mental health among urban adults in England: the moderating effect of household income. Health Place 59, 102200. https://doi.org/10.1016/j.healthplace.2019.102200 Elsevier BV.

Gatys, L.A., Ecker, A.S., Bethge, M., 2016. A neural algorithm of artistic style. J. Vis. 16 (12), 326. https://doi.org/10.1167/16.12.326

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2 (NIPS'14). MIT Press, Cambridge, MA, USA, pp. 2672–2680.

Hoisington, A.J., et al., 2019. Ten questions concerning the built environment and mental health. Build. Environ. 155, 58–69. https://doi.org/10.1016/j.buildenv.2019.03.036 Elsevier BV.

Jaad, A., Abdelghany, K., 2020. Modeling urban growth using video prediction technology: a time-dependent convolutional encoder–decoder architecture. Comput. Aided Civ. Inf. Eng. 35 (5), 430–447. https://doi.org/10.1111/mice.12503 Wiley.

Jean, N., et al., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353 (6301), 790–794. https://doi.org/10.1126/science.aaf7894 United States: American Association for the Advancement of Science.

Li, F., et al., 2009. Built environment and changes in blood pressure in middle aged and older adults. Prev. Med. 48 (3), 237–241. https://doi.org/10.1016/j.ypmed.2009.01.005 Elsevier BV.

Linardatos, P., Papastefanopoulos, V., Kotsiantis, S., 2021. Explainable AI: a review of machine learning interpretability methods. Entropy 23 (1), 18. https://doi.org/10.3390/e23010018 MDPI AG.

Liu, Y., Wu, L., 2016. Geological disaster recognition on optical remote sensing images using deep learning. Prog. Comput. Sci. 91, 566–575. https://doi.org/10.1016/j.procs.2016.07.144 Elsevier BV.

Lopez-Zetina, J., Lee, H., Friis, R., 2006. The link between obesity and the built environment. Evidence from an ecological analysis of obesity and vehicle miles of travel in California. Health Place 12 (4), 656–664. https://doi.org/10.1016/j.healthplace.2005.09.001 Elsevier BV.

Maharana, A., Nsoesie, E.O., 2018. Use of deep learning to examine the association of the built environment with prevalence of neighborhood adult obesity. JAMA Netw. Open 1 (4). https://doi.org/10.1001/jamanetworkopen.2018.1535 United States: American Medical Association.

Marshall, W.E., Piatkowski, D.P., Garrick, N.W., 2014. Community design, street networks, and public health. J. Transp. Health 1 (4), 326–340. https://doi.org/10.1016/j.jth.2014.06.002 Elsevier BV.

May, D., et al., 2009. The impacts of the built environment on health outcomes. Facilities 27 (3–4), 138–151. https://doi.org/10.1108/02632770910933152 Emerald.

McConnell, R., et al., 2006. Traffic, susceptibility, and childhood asthma. Environ. Health Perspect. 114 (5), 766–772. https://doi.org/10.1289/ehp.8594 Environmental Health Perspectives.

Newton, D.W., 2020. Anxious landscapes: correlating the built environment with mental health through deep learning. In: Proceedings of the 40th Annual

Conference of the Association for Computer Aided Design in Architecture (ACADIA): Distributed Proximities, Virtual Conference, October 24–30, 2020. ACADIA, Delaware, pp. 130–139.

Newton, D., 2021. Visualizing deep learning models for urban health analysis. In: Proceedings of the 39th Annual Education and Research in Computer Aided Architectural Design in Europe (eCAADe) Conference: Towards a New Configurable Architecture, Faculty of Technical Sciences, Novi Sad, Serbia, September 8–10, 2021. vol. 1. 527, 536.

Nguyen, A., Yosinski, J., Clune, J., 2019. Understanding neural networks via feature visualization: A survey. In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L., Müller, K.R. (Eds.), Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science. vol. 11700. Springer, Cham. https://doi.org/10.1007/978-3-030-28954-6_4

Peen, J., et al., 2010. The current status of urban-rural differences in psychiatric disorders. Acta Psychiatr. Scand. 121 (2), 84–93. https://doi.org/10.1111/j.1600-0447.2009.01438.x Netherlands.

Piaggesi, S., et al., 2019. Predicting city poverty using satellite imagery. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE Computer Society, Italy. Available at: http://ieeexplore.ieee.org/xpl/conferences.jsp

Rautio, N., et al., 2018. Living environment and its relationship to depressive mood: a systematic review. Int. J. Soc. Psychiatry 64 (1), 92–103. https://doi.org/10.1177/0020764017744582 SAGE Publications.

Renalds, A., Smith, T.H., Hale, P.J., 2010. A systematic review of built environment and health. Fam. Community Health 33 (1), 68–78. https://doi.org/10.1097/fch.0b013e3181c4e2e5 Ovid Technologies (Wolters Kluwer Health).

Robinson, C., Hohman, F., Dilkina, B., 2017. A deep learning approach for population estimation from satellite imagery. In: Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities, GeoHumanities 2017. United States: Association for Computing Machinery, Inc., https://doi.org/10.1145/3149858.3149863

Simonyan, K., Zisserman, A., 2016. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv.org. Available at: https://arxiv.org/abs/1409.1556

Suel, E., et al., 2019. Measuring social, environmental and health inequalities using deep learning and street imagery. Sci. Rep. 9 (1). https://doi.org/10.1038/s41598-019-42036-w Springer Science and Business Media LLC.

Tsagkatakis, G., et al., 2019. Survey of deep-learning approaches for remote sensing observation enhancement. Sensors 19 (18), 3929. https://doi.org/10.3390/s19183929 MDPI AG.

United Nations, Department of Economic and Social Affairs, Population Division, 2019. Population Division. In: World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). United Nations, New York.

WHO, 2016. World Health Statistics 2016: Monitoring Health for the SDGs Sustainable Development Goals. World Health Organization.

Yuchi, W., et al., 2020. Road proximity, air pollution, noise, green space and neurologic disease incidence: a population-based cohort study. Environ. Health 19 (1). https://doi.org/10.1186/s12940-020-0565-4 Springer Science and Business Media LLC.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). United States: Springer Verlag., https://doi.org/10.1007/978-3-319-10590-1_53

Zhang, C., et al., 2019. Joint Deep Learning for land cover and land use classification. Remote Sens. Environ. 221, 173–187. https://doi.org/10.1016/j.rse.2018.11.014 Elsevier BV.

Zhao, H., et al., 2019. Unsupervised deep learning to explore streetscape factors associated with urban cyclist safety. Smart Innov. Syst. Technol. https://doi.org/10.1007/978-981-13-8683-1_16 Australia: Springer Science and Business Media Deutschland GmbH.