Journal of Environmental Management 248 (2019) 109253

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

# Solving visual pollution with deep learning: A new nexus in environmental management

Nahian Ahmed[a], M. Nazmul Islam[a], Ahmad Saraf Tuba[a], M.R.C. Mahdy[a,b], Mohammad Sujauddin[c,*]

[a] Department of Electrical and Computer Engineering, North South University, Bashundhara, Dhaka 1229, Bangladesh
[b] Pi Labs Bangladesh LTD, ARA Bhaban,39, Kazi Nazrul Islam Avenue, Kawran Bazar, Dhaka 1215, Bangladesh
[c] Department of Environmental Science and Management, North South University, Bashundhara, Dhaka 1229, Bangladesh

ARTICLE INFO

*Keywords:*
Visual pollution
Deep learning
Convolutional neural network
Image recognition
Pollutant classification
Environmental management

ABSTRACT

Visual pollution is a relatively new concern amidst the existing plethora of mainstream environmental pollution, recommending the necessity for research to conceptualize, formalize, quantify and assess it from different dimensions. The purpose of this study is to create a new field of automated visual pollutant classification, harnessing the technological prowess of the 21$^{st}$ century for applications in environmental management. From the wide range of visual pollutants, four categories have been considered viz. (i) billboards and signage, (ii) telephone and communication wires, (iii) network and communication towers and (iv) street litter. The deep learning model used in this study simulates the human learning experience in the context of image recognition for visual pollutant classification by training and testing a convolutional neural network with several layers of artificial neurons. Data augmentation using image processing techniques and a train-test split ratio of 80:20 have been used. Training accuracy of 95% and validation accuracy of 85% have been achieved by the deep learning model. The results indicate that the upper limit of accuracy i.e. the asymptote, depends on the dataset size for this type of task. This study has several applications in environmental management. For example, the deployment of the trained model for processing of video/live footage from smartphone applications, closed-circuit television and drones/unmanned aerial vehicles can be applied for both the removal and management of visual pollutants in the natural and built environment. Furthermore, generating the 'visual pollution score/index' of urban regions such as towns and cities will create a new 'metric/indicator' in the field of urban environmental management.

## 1. Introduction

As the built environment of the Earth's surface expands, humans are occupying more land than ever, and cities and towns in developed and developing nations are getting littered with unwanted and unpleasant visual objects; objects that have been termed as 'visual pollutants' (for further details, see section 2). Visual pollution includes not just advertisements, signage, and littered wastes, but any element in the landscape, both indoor and outdoor, that is a misfit for the place and results in an unpleasant, offensive sight (Nagle, 2009). Although the concern against visual pollution is recognized, due to its subjective nature, the problem remains on how to determine a visual pollutant, since what is a visual pollutant to one can be, as extreme as, a beautiful sight to another. Even if an element is determined as a visual pollutant, forces of politics and the capitalist economy work as a barrier towards a visual pollutant-free environment.

Understanding the depth of the possible issues that can be caused by visual pollutants, timely and efficient detection of visual pollutants becomes very important. As of now, the time consuming and expensive method of manual data collection is the sole option for visual pollution research. There is an increasing need for the automation of the aforementioned method of data collection and analysis. Deep learning (for further details see section 4), being the state-of-the-art solution for image recognition, can be used as an efficient alternative for visual pollution data collection. As of now, there is no technical or scientific approach to visual pollution detection and classification through deep learning.

As such, this study, the very first of its kind, will open the window to a new field emphasizing on the technical and scientific aspects of visual pollution, its classifications, and its machine-based (automated)

[*] Corresponding author.
*E-mail addresses:* mohammad.sujauddin@gmail.com, mohammad.sujauddin@northsouth.edu (M. Sujauddin).

https://doi.org/10.1016/j.jenvman.2019.07.024
Received 20 March 2019; Received in revised form 12 June 2019; Accepted 7 July 2019
Available online 12 July 2019

detection processes with applications in the broader spectrum of environmental management. Apart from environmental management, in the future, large scale deployment of the system will aid in deriving visual pollution statistics and metrics across the globe to be used in public health, urban planning, legislative research etc.

## 2. Visual pollution

Portella (2016), in her book “Visual Pollution: Advertising, Signage and Environmental Quality” explored the available literature on the topic. Researchers initially defined visual pollution as the degradation of the “visual quality” of places by advertisements and signage (Ashihara, 1983; Nasar, 1992; Passini, 1992; Cullen, 2000). Scientific literature exists focusing on the damaging effect of the uncontrolled installation of commercial advertisements, mainly billboards, and signage on the landscape, especially that of the scenic and historic sites (Ashihara, 1983; Nasar, 1992; Passini, 1992; Cullen, 2000). Later on, the meaning of the term was expanded to include not just advertisements and signage, but any element in the landscape that is a misfit for the place and results in an unpleasant, offensive sight (Nagle, 2009). Based on this new definition, further literature was published tackling the topic, such as cell phone towers (Nagle, 2009) and wind turbines (Jensen et al., 2014) as “visual pollutants”. Issues of internal and external architecture of buildings and other infrastructures, and their planning were then included under visual pollution (Sumartono, 2009) and research was conducted using time series data that relates the rise of visual pollutants with the development of a city over time. There has been some focus on the management of visual pollutants via software such as GIS (Chmielewski et al., 2016, 2018). For this study, out of all the visual pollutants, we have considered only four major ones, viz. (i) billboards and signage, (ii) telephone and communication wires, (iii) network and communication towers and (iv) street litter (for details, see section 3).

The presence of designing flaws in structures, such as buildings, transportation systems, malls, billboards etc., is a common cause for visual pollution. According to Sumartono (2009), “many apparently properly designed structures take into account functional requirements but not the non-functional ones”. Flaws in the interior design of a structure, such as the color contrast, use of other decorative elements etc. can turn even an apparently well-designed structure into a visual pollutant. In this regard, the use of inappropriate color combination and design can degrade an otherwise good advertisement into a visual pollutant. Some have even considered a disorganized and untidy home as a visual pollutant for its residents.

As in the case of any environment degrading agent, visual pollution results from a “lack of education and culture” (Yilmaz and Sagsöz, 2011) among the common people and especially the governing bodies e.g. lawmakers and law implementers. Due to the lack of awareness and recognition of the adverse psychological, physical, socio-cultural, and economic impacts of visual pollution, the governing bodies allow the existence and growth of such visual pollutants and the mass people, being unaware of its adverse consequences on themselves, do not protest against it. Visual pollutants can also be considered as interferences in achieving/maintaining aestheticism (Mohammadi-Mehr et al., 2018). In addition to this direct effect, lack of awareness also indirectly drives the growth of visual pollutants through unnecessary excessive consumption. According to Yilmaz and Sagsöz (2011), “visual pollution is a result of oversized and unjustified consumption”. As direct evidence, the authors mention the drastic negative change of the downtowns of Turkey due to “tourist commerce proliferation”. Such consumption behaviors are in fact encouraged by today's capitalist society.

It is a fact that man's environment is an indicator of her/his quality of life (Voronych, 2013). Research on the effect of visual pollution on human physiology and psychology have shown that an absence of visual pollutant can reduce the feeling of pain by increasing the secretion of cortisone in the body. Visual pollutant free places have proved to give people a sense of belonging, respect, and pride (Nagle, 2009; Jensen et al., 2014). A visual pollutant-free environment was also proven to increase the social and overall quality of life of the people in that area significantly (Voronych, 2013; Elena et al., 2012). Some commercial gains are also inherent – for example, most of the developed countries are increasingly becoming tourist attractions. This offsets the loss caused by the removal/restriction on outdoor advertisements and other visual pollutants (Elena et al., 2012). Visual pollutants, such as bright light affects insects, disturbing their movement patterns. Once the insects' movement is restricted/hampered, the avian species can no longer have insects as their prey, and this effect continues throughout the food chain, ultimately affecting the humans and the functioning of the whole ecosystem (Elena et al., 2012).

Although a globally accepted standard is yet to be made, according to Portella (2016), there can be a general guideline for ensuring the quality of the visual sphere based on the common views of the users worldwide. But there also needs to be specific guidelines for each and every place/country, since users’ views differ based on their culture/background. At both the national and international level, special regulatory commissions should be formed who will review every new development that will take place for potential visually polluting agents (Elena et al., 2012), and permit only those that abide by the given standards. Many cities have adopted the idea of complete banning of outdoor advertisements. However, it had not been appreciated by all. Relocating the legal structures, in compliance with the zoning laws, have been viewed as a better option. Another possible measure can be to declare certain places as scenic areas, similar to protected areas, where the presence of any potential visual pollutant will be prohibited (Nagle, 2009). Yet, the best measure will be to eradicate visual pollution from the root through creating awareness among the mass people.

## 3. Characterizing the classes of visual pollutants considered in this study[1]

Visual pollutants vary from country to country depending on race, religion, social and economic structure etc. However, this study is not country specific. Rather, the visual pollutants from Bangladesh have been considered, because Bangladesh as a developing country faces serious urban environmental management issues, where there is a pressing demand of a holistic visual pollutant management approach. This approach can be applied for all developing countries in the world by showing Bangladesh as a use case. Among the vast array of visual pollutants discussed in the previous section, four particular visual pollutants have been considered for this study:

### 3.1. *Billboards and signage*

Billboards and signage refer to artificial, usually planar objects placed in the natural and built environment i.e., urban, suburban and city centers, primarily for the purpose of advertisement (Portella, 2016). Billboards and signage were previously an issue that only western countries were encountering. However, in the 21st century, the emergence of developing countries and their increasing hunger for a more 'first world' lifestyle is reflected in the form of these developing countries facing the same issues that first world countries have once faced. Bangladesh could not escape from the devastating negative effects of modernization in the form of excessive use of billboards and signage, instigated by investors to facilitate the consumers' need of being always up to date about the latest products. Certain types of billboards appear in massive numbers during the election fever funded by political campaigns and during festivals as a form of aggressive and

[1] Figs. 1, 2, 3 and 4 (the sources of the images are provided in Supplementary Material 1) show that images collected through web scraping are concurrent with human perception of visual pollution.

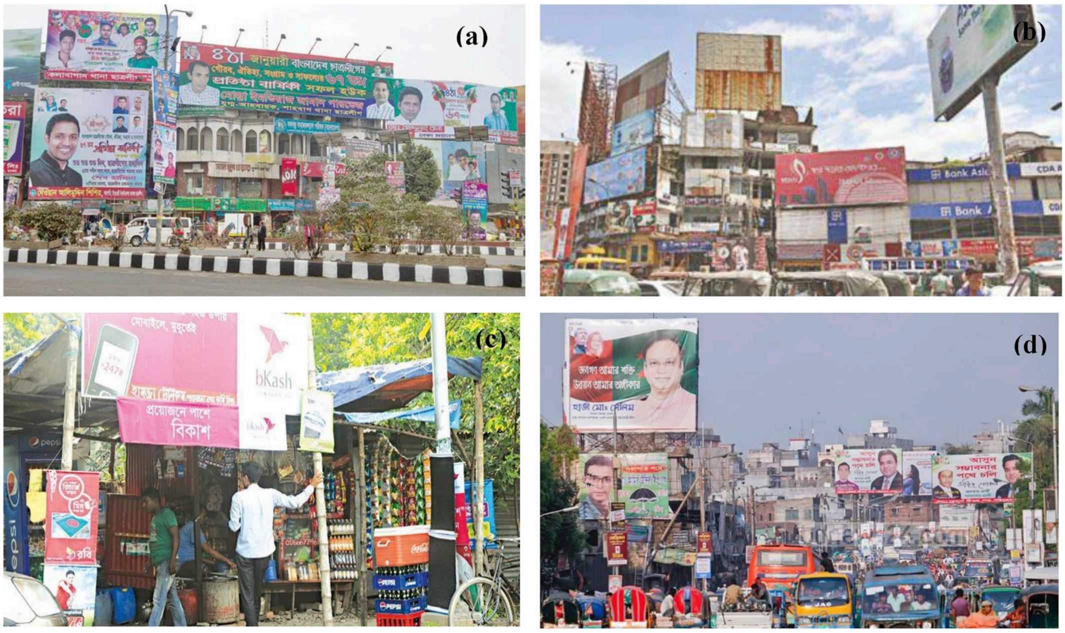


**Fig. 1.** Sample images of the billboards and signage class used for training and testing.

violent product advertisements. In Southeast Asia, local public administration has little to no control over what is built or assembled in public spaces and do not know what is displayed and where it is displayed (Jana and De, 2015) as seen in Fig. 1. For example, advertisements for school coaching centers, room-mate vacancies, housing for sale etc. in the form of billboards and signage can be found in the most unconventional of places such as nailed onto trees, pasted on motorized vehicles, glued on apartment doors etc.

### 3.2. *Telephone and communication wires*

Communication, telephone, and electrical wires are integral part of the modern world, bringing the whole planet just a dial away, implying that anybody can connect to anyone residing anywhere. However, the mismanagement of such entities seems to appear as creeping plants or snakes in the 'modern urban jungle'. These cables and wires do not cause visual pollution until they are tangled in an unorganized or ill-arranged way (Jana and De, 2015). First world countries do not face the problem of visual pollution from these cables and wires since they have moved these cables underground decades ago. Whereas the third world countries like Bangladesh, India, Pakistan etc. are still puzzled about what to do with this massive problem of cables and wires lurking in the proximity of poles and sometimes over the head of the pedestrians. In a country like Bangladesh, these cables have become a threat to pedestrians' life as most of them are tangled with the electrical wires, as well as creating a huge visual impact on the eyes of visitors and pedestrians of urban and suburban city centers as they block the natural view of the aesthetic features related to the history of those cities (Fig. 2).

### 3.3. *Network and communication towers*

Network and communication towers, also known as cell phone towers, refers to structural objects planted in the natural environment (e.g. in rural areas of Bangladesh) or built environment (e.g. rooftop of high-rise buildings in urban or suburban city centers). Cell phone towers came to the scene of the 21$^{st}$ century after communication systems across the world were (and still are) competing to become wireless, facilitating the use of wireless handheld devices such as smartphones. As the cell phone fever arrived in Bangladesh in the early 1990s, telecom operators, investors, and manufacturers took the advantage and hacked the growth of their business by deploying towers at a remarkable rate. Now, Bangladesh has cell phone towers almost everywhere, hindering the view of a picturesque landscape, a forest, hill, greeneries or even the vast view from the rooftop in urban areas (Fig. 3) concurring with Nagle (2009) who has stated “view is everything and a tower kills the view”.

### 3.4. *Street litter*

First world countries have accounted for street litter decades ago, having proper plans and setup for waste disposal requiring investments of vast sum of funds. These countries have also concentrated on reusable products from solid waste, contributing to reduction of street waste visual pollution though many old towns of first world countries still suffer from this problem. Inhabitants of South Asian third world countries like Bangladesh, India, and Pakistan, lack the awareness and mindset that waste should be handled in a proper way leading to disposal on urban streets and open public spaces, forming ‘garbage heaps’ (Fig. 4). But the most devastating effect is that the population has gotten used to such an environment which can result in a character-changing impact on the community as a whole, further paving the way for the loss of quality of life (Jana and De, 2015).

## 4. Deep learning and its applications in environmental management

In the 'Information Age', communication and transmission of energy has become an integral and inseparable part of modern life, the technology required for which has brought along a vast plethora of physical devices/structures such as network towers, communication and electric wires etc. Billboards and signage (both digital and painted/printed) was introduced in the 20th century and is currently at its peak in terms of both scale and scope, with no sign of slowing down. Personalized advertisements in social media sites are exemplary in portraying the extent of consumerism in today's age where advertisements have invaded almost every aspect of both public and personal life. Thus, it is apparent that a major proportion of visual pollutants are in fact components or implications of some form of technology.

However, technology itself can be applied in the recording, management and mitigation of visual pollution. Detection and classification

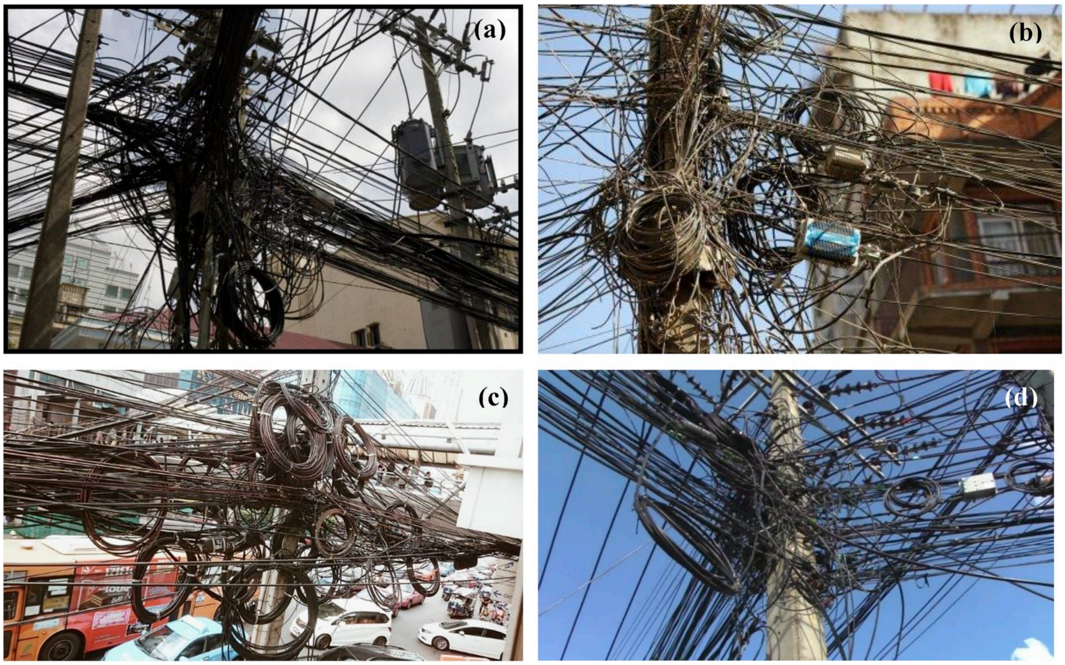


**Fig. 2.** Sample images of telephone and communication wire class used for training and testing.

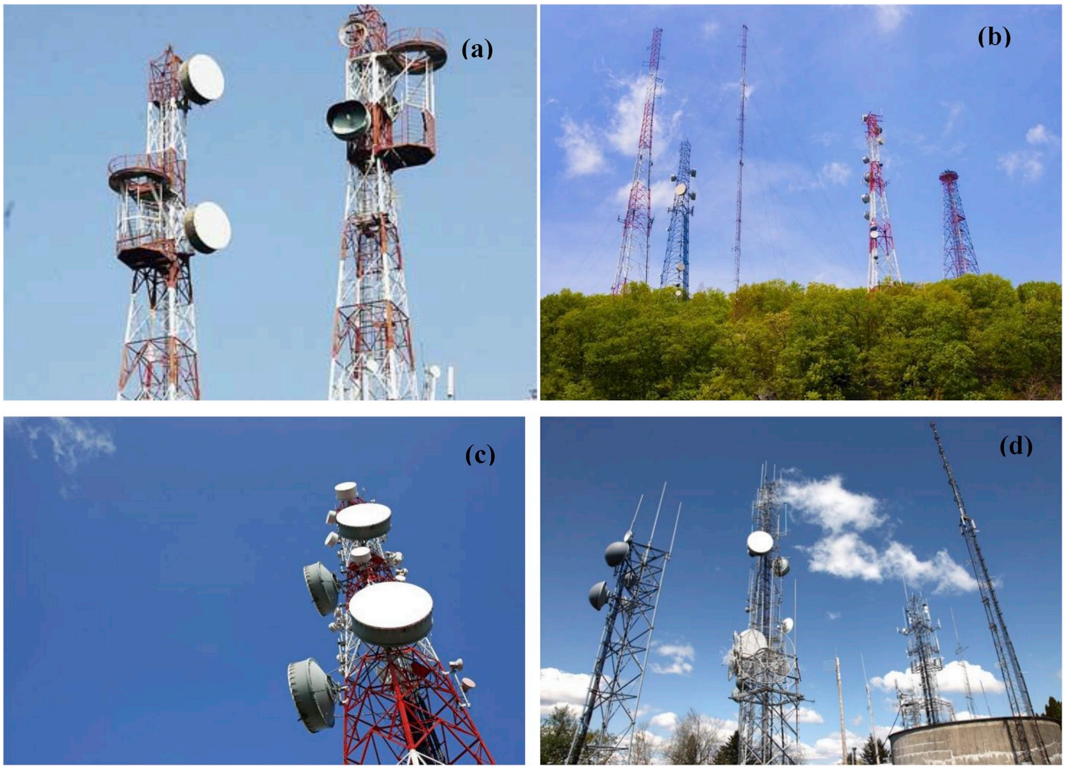


**Fig. 3.** Sample images of network tower class used for training and testing.

of visual pollutants by a human is a trivial task whereas designing and building an automated system that performs the same task is not trivial at all. Images of the natural and built environment are analogous to the visual areas of a person. Visual pollutants (in images) can have infinite combination of shape, color, size etc. which may depend on angle of perception, lighting, geographic location etc. It is impossible to hard code such rules since computation would be very time consuming (if possible, at all) and highly error prone. Thus, a generalized model is necessary to perform the task and requires the involvement of machine learning. Since the human brain is exceptionally successful at performing such tasks, it is beneficiary to model the system after the human brain itself. A mathematical model of the human brain (McCulloch and Pitts, 1943) also known as artificial neural network (ANN) was proposed over half a century ago where the neurons of the human brain are modeled as artificial neurons. Several variants have emerged depending on the specific application of ANNs. ANNs have layer(s) of neurons where the output of one layer is fed as input to the successive layer (analogous to the neurons in the human brain), where

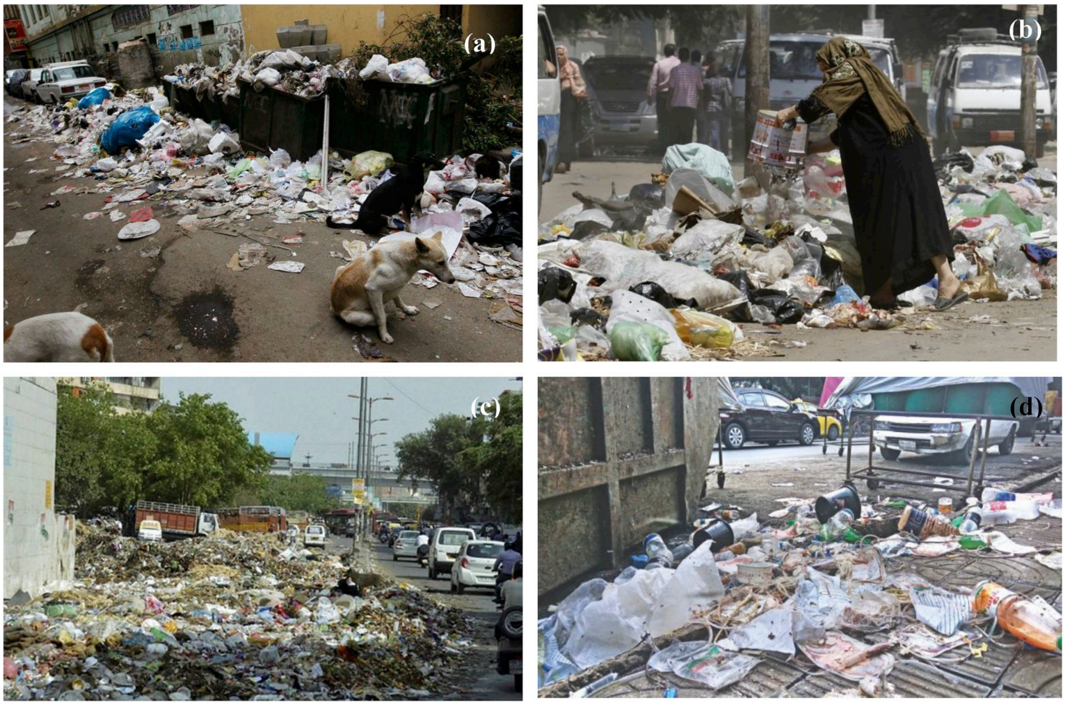


**Fig. 4.** Sample images of street litter class used for training and testing.

the number of layers is correlated to the complexity of the task being performed i.e., a harder task requires more layers of artificial neurons. Deep neural networks (DNN) are variants of ANN, the learning system for which is generally known as 'deep learning'. Thus, the word 'deep' in deep learning refers to the fact that DNNs have several layers of artificial neurons.

Deep learning (LeCun et al., 2015) has become the state-of-the-art for image recognition tasks. Deep learning models have been used for medical image analysis (Litjens et al., 2017), for identification of coral reef species (Villon et al., 2018), for predicting poverty from satellite images (Jean et al., 2016), for satellite image scene classification (Zou et al., 2015), for detection of animals in footage collected from unmanned aerial vehicles (Rey et al., 2017; Kellenberger et al., 2018) etc. and thus hold great promise for being a solution in automating the visually polluting image classification process. Furthermore, wide variety of powerful and sophisticated deep learning models are present in literature (Krizhevsky et al., 2012; He et al., 2015, 2016; Rastegari et al., 2016; Szegedy et al., 2016).

## 5. Materials and methods

### 5.1. Collection of images

Labeled images are necessary for training a deep learning classifier which were obtained via the Google Image Search engine. The human perception of visually polluting entities in the physical environment (captured in digital images) is reflected in the search results. Images showing up in results are linked via keywords such as 'visual pollution'. Thus, there is a direct relationship between the images showing up in search results and images which humans perceive as visually polluting since the images on the web themselves were uploaded by humans who perceived the image as visually polluting. Furthermore, inclusion of keywords such as 'billboard', 'telephone wire' etc. along with 'visual pollution', provides pre-labeled images. The term 'Bangladesh' is also provided to narrow down on region specific images of visual pollutants. However, such specificity has certain drawbacks i.e. lower number of training and testing images (since Bangladesh is a small region compared to the world) and inclusion of images of geographically similar regions i.e. countries of South East Asia such as India, Malaysia, Vietnam etc. High particularity in image collection achieved by providing specialized search terms such as 'Bangladesh visual pollution billboard', limits the number of appropriate/useable images for each class. This method of image data collection, though prone to some error (e.g. 'outlier images'; for further details see section 5.2), is extremely fast and cost effective. The characterization, sources, and effects of the four visual pollutant classes that have been considered in this study are discussed previously in Section 3. Python scripting was used for collecting large amounts of images from the web in batches.

### 5.2. Preprocessing and augmentation

Though majority of images showing up as search results contain visual pollutants when keywords such as 'visual pollution' are supplied, there are some 'outlier images' i.e. irrelevant images showing up in search results which are in fact not visually polluting but show up due to caveats in the inner workings of the search engine algorithm. Manual intervention is necessary in identifying these 'outlier images' and excluding them from training and testing of the model. After removal of irrelevant images, 200 images of each class were obtained leading to 800 total images. Deep learning models outperform traditional machine learning models due to the availability of large amount of data. Where the accuracy of traditional machine learning models plateaus off, accuracy of deep learning models keeps increasing with amount of data.[2] Thus, image augmentation was used for creating several images from each image to increase the amount of data synthetically. Image processing algorithms such as translation, rotation and flip are used for generating augmented images. This also makes the models translation, rotation and flip invariant, meaning that models will accurately identify visual pollutants regardless of orientation of the entity within the image

[2] Due to limitations of funding, manual collection of large number of images was not possible. However, we have utilized the available image resources on the web for the purpose of this study.

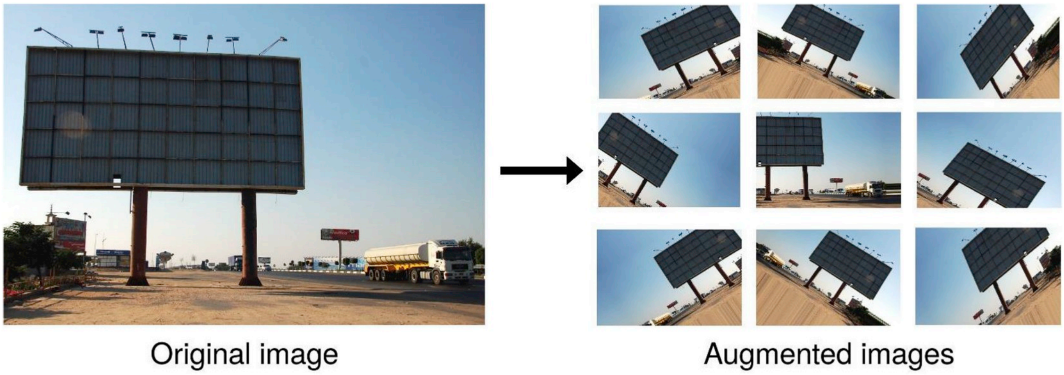


**Fig. 5.** Schematic of augmentation applied on training images. An image of a billboard is used as example; the original image[3] containing the visual pollutant (left) is augmented to create images with different orientations (right).

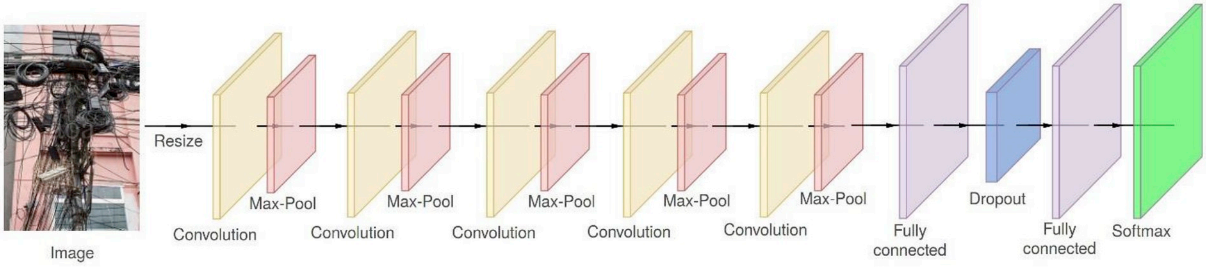


**Fig. 6.** System architecture of the sequential deep learning model for visual pollutant classification.

and/or the angle of perception/position of viewer. A sample image augmentation process is shown in Fig. 5. Since all images are of different sizes and the input layer of the CNN has fixed size, the input dimensions of images need to be equalized. Consequently, images are resized to dimension of 150 × 150 pixels.

### 5.3. System architecture and experimental setup

The system architecture of the CNN used in this study is shown in Fig. 6. After the images are resized (as mentioned in previous subsection), the images are passed through convolution and max-pooling layers 5 times, extracting higher level features from the image as it passes through the sequential deep learning model. Lower level features such as edges in the images are combined to form higher level features such as shapes, in succeeding layers. For the regularization of model weights, a dropout layer with dropout rate of 50% and L2 kernel regularizer is used in the second to last fully connected layer. Without regularization, models tend to overfit to training data, reducing validation accuracy.

All layers within the deep learning model except for the last layer (softmax layer) use the Rectified Linear Unit (ReLU) activation function which is mathematically shown as

$$R(z) = \max(0,\ z)$$

where the output is zero if the input $z$ is negative or else the activation function behaves as an identity activation function in the positive domain.

On the other hand, the last layer uses the softmax activation function for generating the probability of the input images being members of a certain visual pollutant class. The softmax activation function can be shown as

$$\sigma(x_j) = \frac{e^{x_j}}{\sum_i e^{x_i}}$$

where $j$ corresponds to a specific class of visual pollutant and $x_j$ is the value arriving at the corresponding artificial neuron for that class of visual pollutant in the softmax layer. Summation of output of neurons in final layer over $i$ is the denominator and represents the sum of probability output of all four output neurons. Since we have 4 visual pollutant classes the output layer will have 4 neurons, and both $i$ and $j$ will have values between 0 and 3 inclusive (considering that classes are zero indexed).

The training size to testing size ratio is 80:20. Thus, 640 images were used for training and 160 images were used for validation. The models are trained for a total of 50 epochs (see section 5.3 for further details) and batch size of 16 is used leading to 125 iterations in each epoch. Due to the type of the learning problem being a multi-label supervised classification one, the categorical cross entropy loss function was used. The RMSprop optimizer was used for weight optimization of model.

### 5.4. Software used

The Python API (Van Rossum and Drake, 1995) for Keras with TensorFlow (Abadi et al., 2016) back end is used for model training and testing. The ggplot2 package (Wickham, 2016) of R (R Core Team, 2013) is used for visualization of results.

## 6. Results and discussion

### 6.1. How image augmentation is related to the human learning experience?

Image augmentation involves creating new images (augmenting) from an existing image by applying combinations of image processing algorithm (mentioned in section 5.2) as shown in Fig. 5. The augmented images contain the visual pollutant(s) in the original image while changing the orientation of the pollutant(s) in the image. As a result, during the training of the deep learning model, images are provided in which the same visual pollutant(s) have been photographed from different perspectives/viewpoints, aiding the model in generalizing well

[3] See Supplementary material for source of original image.

and achieving higher accuracy. The human learning experience shows that objects in a person's visual areas are often recognized well after it has been observed from different viewpoints and at different orientations. The 'methodological' resemblance between this phenomenon and image augmentation is quite uncanny. However, it is important to note that image augmentation can only finitely and restrictedly mimic the complexities of the human learning experience.

### 6.2. Model accuracy analysis

Passing the entire dataset forward and backward through the deep learning model once is considered as one epoch of training. Since feeding the entire dataset to the model every epoch is computationally challenging, the dataset is passed through the network (both forward pass and backward pass) in batches. The model has an overall cost function i.e. loss function where different values of losses are obtained based on the specific configurations of the weights of the network. Since the data is constant and the weights are variable, optimization of weight parameters is equivalent to model performance optimization. Whereas the optimization of weight parameters is equivalent to finding the permutation of weights which provides a cost/loss which is a minima (preferably global) i.e. where the first derivative is approximately zero and the second derivative is positive, in the cost/loss surface in $\mathbb{R}^n$ ($n$-dimensional space) where $n$ is the total number of weights in the deep learning model. Loss/cost and accuracy are inversely correlated. Accuracy is the percentage of correct classifications made by the deep learning model on the 160 validation images.

Fig. 7 shows the categorical cross-entropy loss and accuracy for the model during the 50 epochs of training. As training of the deep learning model progresses i.e. with successive epochs, the weights are shifted to more 'optimized' values, reducing loss and increasing the accuracy of the model. The training loss is constantly decreasing and training accuracy is constantly increasing. This is because the model is being trained on the same 640 images over and over again and the training accuracy refers to the number of correct classifications the model makes on those 640 images, since the model gets highly 'specialized' at recognizing these images, the loss/cost gets lower and accuracy gets higher. However, the 160 validation images are images that the deep learning model has never previously 'seen', thus the validation loss and accuracy depend on how well the model can generalize to new images rather than keep specializing on images it has been trained on. This explains why validation loss is almost always higher than training loss and validation accuracy is almost always lower than training accuracy.

To summarize the training and testing performance of the model, the 50-epoch training interval is divided into 10 intervals each representing 5 epochs of training. The mean and standard deviation of training loss, validation loss, training accuracy and validation accuracy for those 10 intervals are shown in Table 1. When a specific epoch is validated using specific images that the classifier finds difficult to recognize correctly, the validation loss increases and the validation accuracy falls. As Fig. 7 and Table 1 shows that during epoch 16–20 and epoch 41–45 the loss increases and accuracy decreases sharply due to this reason. However, as training progresses the loss decreases and accuracy increases again as the weight vector maps the inputs with minimal cost/loss. Fig. 7 shows that validation accuracy reaches the highest value of 87% at epoch 46. However, Table 1 shows that the mean accuracy of the interval of epoch 46–50 is 85%. This is because there is some variability of accuracy, which is reflected in the form of the standard deviation of the validation accuracy. With training, the variability of the results of the model decrease i.e. standard deviation of validation accuracy decreases with subsequent training intervals.

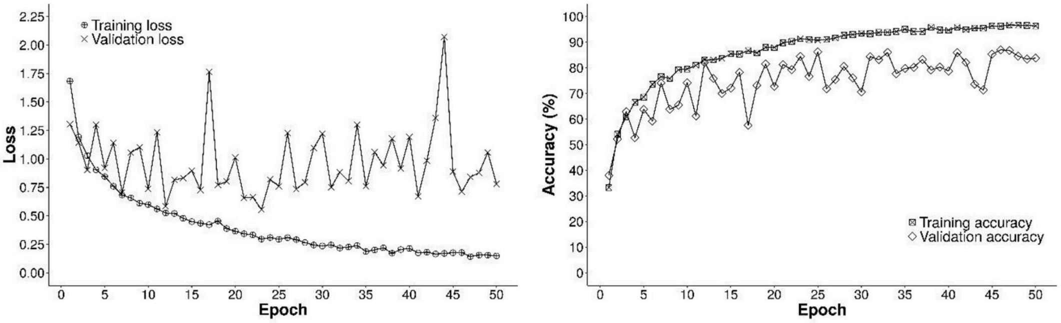
**Table 1**
Statistical summary of model performance metrics segmented intro 5-epoch intervals.

| Epoch | Training loss | | Validation Loss | | Training accuracy (%) | | Validation accuracy (%) | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1–5 | 1.13 | 0.30 | 1.11 | 0.18 | 56.65 | 12.72 | 53.93 | 9.33 |
| 6–10 | 0.66 | 0.06 | 0.95 | 0.19 | 76.93 | 2.21 | 67.34 | 5.91 |
| 11–15 | 0.51 | 0.04 | 0.87 | 0.21 | 83.26 | 1.44 | 72.23 | 6.90 |
| 16–20 | 0.41 | 0.03 | 1.02 | 0.39 | 86.74 | 1.06 | 72.61 | 8.21 |
| 21–25 | 0.31 | 0.02 | 0.69 | 0.09 | 90.63 | 0.58 | 81.55 | 3.46 |
| 26–30 | 0.27 | 0.03 | 1.02 | 0.21 | 92.36 | 0.89 | 74.86 | 3.51 |
| 31–35 | 0.22 | 0.02 | 0.90 | 0.20 | 94.00 | 0.60 | 82.19 | 3.04 |
| 36–40 | 0.20 | 0.02 | 1.06 | 0.11 | 94.69 | 0.64 | 80.30 | 1.58 |
| 41–45 | 0.17 | 0.01 | 1.12 | 0.49 | 95.59 | 0.50 | 79.66 | 6.06 |
| 46–50 | 0.15 | 0.01 | 0.85 | 0.12 | **96.46** | 0.18 | **85.09** | 1.51 |

### 6.3. The relationship among deep learning, human perception and visual pollutant classification

The analogy between deep learning and the human learning process/experience is quite apparent. As a result, there are several similarities which can be observed at all levels of representations in terms of systems and sub-systems of components. For example, the training time required for a model to learn to correctly classify can be quantified in terms of epochs. Considering its counterpart, the 'training time' of the human learning experience, it can be compared to the number of times a human has to see, identify and acknowledge a visual pollutant in the environment before that person can correctly classify a visual pollutant i.e. until specific connections between neurons in the brain have been established, this phenomenon itself is analogous to the weight change concept of deep learning.

The whole concept of feature maps in convolutional neural networks are for extracting higher level features from an image. For



(a) Categorical cross entropy loss

(b) Percentage of images being classified accurately

**Fig. 7.** Model performance metrics monitored throughout training and testing.

example, classifying an image as a member of the billboard and signage class may depend on the presence of a rectangular object in the image. The deep learning model does so by first extracting features such as specific edges in the image, as the image is passed through successive convolution and max-pooling layers, the edges form larger shapes such as rectangles, which in turn determines the output of the softmax layer. The human learning experience also has a similar analogy since humans tend to associate billboards with the shape of a rectangle.

On the other hand, there are several fundamental differences between deep learning and the human learning process. After all, artificial neurons are abstractions/models of the biological neuron. In truth, science provides ‘models’, not the exact implementations, there is an indefinite race to define models that incorporate the finer details, for example, classical physics and quantum physics. Though there is more ‘applicability’ of classical physics in terms of magnitude in today's world, substantial research is dedicated to the field of quantum physics to increase the applicability of that contemporary field. The same can be considered for artificial neurons and artificial neural networks. As a consequence, there is significant research on modelling of the biological neuron (Hopfield, 1984; Mahowald and Douglas, 1991; Izhikevich, 2003; Oprisan et al., 2004; Conte et al., 2006). There are many similarities between the biological neuron and the artificial neuron. For example, the nucleus/cell body of the biological neuron is represented as the mathematical operations that occur in the artificial neuron i.e. summation and application of activation function. The dendrites represent the inputs and the axon represents the output(s).

However, it is also true that artificial neurons are abstractions of a biological neuron which require certain assumptions to be made, introducing inherent error due to modelling limitations. Currently, computer hardware is getting specialized for deep learning i.e. the use of graphics processing units (GPUs), which was used in this study. Even more specialized hardware for deep learning is now present i.e. tensor processing units (TPUs). There has been (and will be) an exponential increase in computational power, continuing Moore's law (Schaller, 1997). Tasks such as more realistic/detailed modelling of the biological neuron which was previously computationally unfeasible is currently possible (Gleeson et al., 2010; Marder and Taylor, 2011). More realistic mathematical modelling of the biological neuron will change the structures of existing state-of-the-art neurons and neural network architectures. Thus, both artificial neurons and artificial neural networks need to better mimic biological neurons and neural networks while maintaining computation cost realistic.

The concept of a ‘universal visual pollutant classifier’ is governed by four issues viz., i) the number of different visual pollutant classes, since having large number of classes will increase the state space of the model, having potentially negative effect on the accuracy unless dataset size is increased both quantitatively (more images of each class) and qualitatively (more classes of images), ii) the presence of a hierarchical system of categorization of visual pollutants i.e. separate classes for election poster and billboards instead of a single generalized class, iii) the intra-class homogeneity of the visual pollutant classes which represents how similar images of the same visual pollutant category are and iv) the inter-class heterogeneity, which represents how distinct two images of two different visual pollutant classes are.

### 6.4. *Scaling the mountain: the asymptote of accuracy*

Due to the asymptotic and exponential nature of the accuracy function, as seen in Fig. 7, going from 40% to 50% is easier than going from 50% to 60%. Fig. 7 shows that the validation accuracy increased by approximately 14% from epoch 1 to epoch 2 and 10% from epoch 2 to epoch 3. This shows that the overall validation accuracy increase reduces with subsequent epochs. The validation accuracy increased by less than 1% in epoch 49 to epoch 50, confirming the asymptotic and exponential nature of validation accuracy. As a result, increasing the baseline validation accuracy, say from, 80%–90%, would require several magnitudes more images. It also depends on the random splitting of the entire dataset into training and testing samples since validation accuracy is largely affected by the images in the testing/validation data.

## 7. Applicability of this research in environmental management

This study introduces an interdisciplinary field of automated visual pollution detection which sits at the intersection of visual pollution, aestheticism, human perception, deep learning/machine learning and environmental management with massive applicability. There are several potential applications of the deep learning based visual pollutant classifier developed in this study, especially in the context of environmental management.

*Video and live footage analysis:* Video/live footage are sequences of images being played at a constant rate i.e. with regular time intervals between each image. Since the deep learning model developed in this study can classify visual pollutant in images, it can also be used for the same task on video/live footage. Each frame of the video/live footage is extracted and fed to the trained model as input which recognizes the visual pollutants in the frame/image and records the result in a database. Footage from surveillance cameras installed in urban areas can be processed using the model to detect and classify visual pollutants.

*Cloud-based server deployment:* The model deployed on a server would make it accessible to anyone with an internet connection, even from remote regions across the globe. Furthermore, deployment on the cloud would mean that several instances of the model could be run simultaneously by different parties.

*Smartphone application development:* The model can be used for the development of camera-based smartphone applications which could be used for both image collection in the future as well as for detection purposes. There are approximately 2.71 billion smartphone users across the globe in 2019 which will rise to 2.87 billion in 2020 (eMarketer, n.d.). Thus, the potential user base for a deep learning based visual pollutant classifier on a smartphone application is massive.

*Equipment on drones and unmanned aerial vehicles (UAVs):* Aerial devices equipped with a visual pollutant classifier can be used for obtaining images of visual pollutants as well as the geographic coordinates through GPS.

*Visual pollutant removal and management:* Once the location of the visual pollutants have been obtained through smartphone applications and drones/UAVs, manual intervention can be applied more effectively for both the removal and management of visual pollutants in the natural and built environment. For example, the model could be used for locating street litter which needs removing, aiding in solid waste management.

*Visual pollution index generation:* The information retrieved from smartphone and drones/UAVs equipped with the model can be used for generating a visual pollution index for geographic regions such as a city, town or country. This would be a very important tool for urban planners and professionals in the field of urban environmental management. The visual pollution index could be used for both evaluation of the visual aestheticism of a geographic region as well as comparison between the visual aestheticism of different geographic regions. Geolocations of automatically identified visual pollutants can be used to calculate the ‘visual pollutant density’ which is a direct contributor to the visual pollutant index.

*Methodological replicability:* The methodology developed in this study can be used for training deep learning models (and other machine learning models) specialized for specific visual pollutants of a different part of the world as well as for larger datasets with more visual pollutant classes.

*Transfer learning:* As the results indicate, the model achieves approximately 85% validation accuracy. This relatively high validation accuracy (even though being severely restricted by dataset of only 800 images) hint towards the huge potential of this model to be used for transfer learning. Thus, the pre-trained model can directly be used for classifying new images containing visual pollutants.

*Industrial scale products:* This study instigates the development of both hardware and software products specialized for visual pollutant classification. For example, development of specialized GPU architectures for model training and execution of visual pollutant classifiers can massively reduce training and detection time needed for the systems. Software products can be developed for image and video analysis. With significant awareness about visual pollution and aestheticism across the globe, these products will become an integral part of environmental management.

## 8. Conclusion

The results of this study show that even an abstract simulation of a component of the human brain i.e. through deep learning, is capable of achieving validation accuracy of 85% and training accuracy of 95% for the highly complex task of visual pollutant classification. Thus, more realistic simulations through deep learning models specialized for visual pollutant classification will increase both the validation accuracy achieved as well as the overall applicability of the study. Simulations of the human learning experience such as image augmentation plays an important role in deep learning, especially if the data set size is constrained. The automated data collection method has several drawbacks. There is a limit to the number of appropriate images that can be obtained for a specific visual pollutant. Whereas, manual collection of images can be used to created datasets of millions of images.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2019.07.024.

## References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al., 2016. Tensorflow: a system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), pp. 265–283.

Ashihara, Y., 1983. The Aesthetic Townscape. MIT Press, Cambridge.

Chmielewski, S., Lee, D.J., Tompalski, P., Chmielewski, T.J., Wężyk, P., 2016. Measuring visual pollution by outdoor advertisements in an urban street using intervisibilty analysis and public surveys. Int. J. Geogr. Inf. Sci. 30 (4), 801–818.

Chmielewski, S., Samulowska, M., Lupa, M., Lee, D.J., Zagajewski, B., 2018. Citizen science and WebGIS for outdoor advertisement visual pollution assessment. Comput. Environ. Urban Syst. 67, 97–109.

Conte, E., Pierri, G., Federici, A., Mendolicchio, L., Zbilut, J.P., 2006. A model of biological neuron with terminal chaos and quantum-like features. Chaos, Solit. Fractals 30 (4), 774–780.

Cullen, G., 2000. The Concise Townscape. Architectural Press, Oxford.

Elena, E., Cristian, M., Suzana, P., 2012. Visual pollution: a new axiological dimension of marketing? Ann. Fac. Econ. 1 (2), 820–826.

eMarketer. (n.d.). Number of smartphone users worldwide from 2014 to 2020 (in billions). In Statista - The Statistics Portal. Retrieved February 26, 2019, from https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/.

Gleeson, P., Crook, S., Cannon, R.C., Hines, M.L., Billings, G.O., Farinella, M., et al., 2010. NeuroML: a language for describing data driven models of neurons and networks with a high degree of biological detail. PLoS Comput. Biol. 6 (6), e1000815.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Hopfield, J.J., 1984. Neurons with graded response have collective computational properties like those of two-state neurons. Proc. Natl. Acad. Sci. Unit. States Am. 81 (10), 3088–3092.

Izhikevich, E.M., 2003. Simple model of spiking neurons. IEEE Trans. Neural Netw. 14 (6), 1569–1572.

Jana, M.K., De, T., 2015. Visual pollution can have a deep degrading effect on urban and suburban community: a study in few places of bengal, India, with special reference to unorganized billboards. Eur. Sci. J.,ESJ 11 (10).

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353 (6301), 790–794.

Jensen, C.U., Panduro, T.E., Lundhede, T.H., 2014. The vindication of Don Quixote: the impact of noise and visual pollution from wind turbines. Land Econ. 90 (4), 668–682.

Kellenberger, B., Marcos, D., Tuia, D., 2018. Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning. Remote Sens. Environ. 216, 139–153.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., et al., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Mahowald, M., Douglas, R., 1991. A silicon neuron. Nature 354 (6354), 515.

Marder, E., Taylor, A.L., 2011. Multiple models to capture the variability in biological neurons and networks. Nat. Neurosci. 14 (2), 133.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5 (4), 115–133.

Mohammadi-Mehr, S., Bijani, M., Abbasi, E., 2018. Factors affecting the aesthetic behavior of villagers towards the natural environment: the case of kermanshah province, Iran. J. Agric. Sci. Technol. A 20 (7), 1353–1367.

Nagle, J.C., 2009. Cell phone towers as visual pollution. Notre Dame JL Ethics & Pub. Pol'y 23, 537.

Nasar, J.L. (Ed.), 1992. Environmental Aesthetics: Theory, Research, and Application. Cambridge University Press.

Oprisan, S.A., Prinz, A.A., Canavier, C.C., 2004. Phase resetting and phase locking in hybrid circuits of one model and one biological neuron. Biophys. J. 87 (4), 2283–2298.

Passini, R., 1992. Wayfinding: People, Signs, and Architecture. McGraw-Hill Ryerson.

Portella, A., 2016. Visual Pollution: Advertising, Signage and Environmental Quality. Routledge.

R Core Team, 2013. R: A Language and Environment for Statistical Computing.

Rastegari, M., Ordonez, V., Redmon, J., Farhadi, A., 2016. October). Xnor-net: imagenet classification using binary convolutional neural networks. In: European Conference on Computer Vision. Springer, Cham, pp. 525–542.

Rey, N., Volpi, M., Joost, S., Tuia, D., 2017. Detecting animals in african savanna with UAVs and the crowds. Remote Sens. Environ. 200, 341–351.

Schaller, R.R., 1997. Moore's law: past, present and future. IEEE Spectr. 34 (6), 52–59.

Sumartono, S., 2009. Visual pollution in the context of conflicting design requirements. J. Vis. Art Des. 3 (2), 187–196.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.

Van Rossum, G., Drake Jr., F.L., 1995. Python Reference Manual. Centrum voor Wiskunde en Informatica, Amsterdam.

Villon, S., Mouillot, D., Chaumont, M., Darling, E.S., Subsol, G., Claverie, T., Villéger, S., 2018. A Deep learning method for accurate and fast identification of coral reef fishes in underwater images. Ecol. Inf. 48, 238–244.

Voronych, Y., 2013. Visual Pollution of Urban Space in Lviv. Przestrzeń I Forma.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer.

Yilmaz, D., Sagsöz, A., 2011. In the context of visual pollution: effects to trabzon city center Silhoutte. Asian Soc. Sci. 7 (5), 98.

Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015. Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. 12 (11), 2321–2325.