Case Study

Deep Learning for Critical Infrastructure Resilience

Kevin Dick[1]; Luke Russell[2]; Yasmina Souley Dosso[3]; Felix Kwamena, Ph.D.[4]; and James R. Green, Ph.D., P.Eng.[5]

Abstract: Ensuring the resiliency of critical infrastructures is essential in modern society, but much of the deployed infrastructure has yet to fully leverage modern technical developments. This paper intersects two unique fields—deep learning and critical infrastructure protection—and illustrates how deep learning can improve resiliency within the electricity sector. Machine vision is the combination of machine intelligence, or computer systems automatically learning patterns from exemplar data, and image analysis, or objects of interest being automatically segmented and identified from video image data. This technology has the potential to automate threat assessments in the context of securing critical infrastructures. Rather than traditional reactionary approaches, we present here a method of leveraging deep learning for the detection of threats to critical infrastructures before failures occur. This paper discusses the state-of-the-art in deep learning for creating machine vision systems, and the concepts are applied to increase the resiliency of critical infrastructures. The intersection between machine vision and critical infrastructures is discussed, as are key benefits and challenges of invoking such an approach, and examples within several fields of critical infrastructures are presented. Automated inspection of the power infrastructure using vehicle-mounted video acquisition equipment is explored, and a proof-of-concept implementation of a deep convolutional neural network is developed, achieving 95.5% accuracy in distinguishing power-related infrastructures within images largely typical of rural settings. These preliminary results show promise in the application of deep learning and machine vision to protecting critical infrastructures through preventative maintenance. **DOI: 10.1061/(ASCE)IS.1943-555X.0000477.**

Introduction

High-tech innovation is poised to transform the energy industry as smarter management of complex systems, data analytics, and automation is developed to address energy infrastructure security and resilience (Victor and Yanosek 2017). Engineering solutions have the potential to improve the resilience of critical infrastructures. In particular, given an increasing supply of video and photographical information, the concepts of machine vision and deep learning can be leveraged to improve the resilience of critical infrastructures by processing image data to highlight anomalies before a failure occurs.

Nation's Critical infrastructures

The importance of critical infrastructures is recognized internationally. For example, in Canada, a critical infrastructure is defined as the "processes, systems, facilities, technologies, networks, assets and services essential to the health, safety, security or economic well-being" of people, as well as the effective functioning of government (Government of Canada 2017); similar definitions are used across many countries of the Organisation for Economic Co-operation and Development. Nations and private owner-operators are faced with the complex obligation to maintain and protect critical infrastructure assets. Such an obligation constitutes a logistical challenge because many critical infrastructure elements are often owned by the private sector and are regulated by governments. Critical infrastructure protection (CIP) programs typically have three aims. The first is to apply a combination of security measures to increase the resiliency of critical infrastructures against both intentional and accidental incidents. Second, CIP programs establish business continuity procedures for essential services to minimize disruptions from such incidents. Finally, CIP programs develop appropriate and sufficient emergency response procedures in the event of disruptions or disasters (Public Safety Canada 2015). Common to these three aims is the concept of *resilience*, which can be broadly defined as "the intrinsic ability of a system to mitigate disturbances to regain or sustain required operations through both expected and unexpected conditions." Disruptions to critical infrastructures could generate wide-ranging economic and societal impacts, potentially resulting in "catastrophic loss of life, adverse economic effects and significant harm to public confidence" (Government of Canada 2017). Identifying and implementing technological solutions to improve CIP promises to reduce disruptions through increased systematic resilience.

[1]Ph.D. Candidate, Dept. of Systems and Computer Engineering, Carleton Univ., Ottawa, ON, Canada K1S 5B6. Email: kevin.dick@carleton.ca

[2]Researcher, Dept. of Systems and Computer Engineering, Carleton Univ., Ottawa, ON, Canada K1S 5B6. Email: luke.russell@carleton.ca

[3]Ph.D. Candidate, Dept. of Systems and Computer Engineering, Carleton Univ., Ottawa, ON, Canada K1S 5B6. Email: yasmina.souleydosso@carleton.ca

[4]Adjunct Professor, Dept. of Systems and Computer Engineering, Carleton Univ., Ottawa, ON, Canada K1S 5B6. Email: felix.kwamena@carleton.ca

[5]Associate Professor, Dept. of Systems and Computer Engineering, Carleton Univ., Ottawa, ON, Canada K1S 5B6 (corresponding author). ORCID: https://orcid.org/0000-0002-6039-2355. Email: jrgreen@sce.carleton.ca

Note. This manuscript was submitted on March 21, 2018; approved on September 13, 2018; published online on January 31, 2019. Discussion period open until June 30, 2019; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, © ASCE, ISSN 1076-0342.

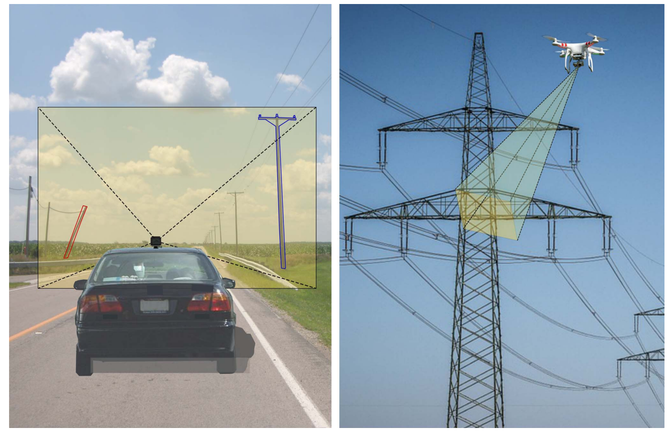
Resilience Innovation for Preventative Maintenance

Technological innovation is a fundamental component to creating CIP frameworks and procedures that are required to ensure recovery and implementation of integrated systems for a resilient country to function (Community and Regional Resilience Institute 2013). The authors define *resilience innovation* as "the process of developing robust systems or improving upon existing infrastructure such that they are increasingly resilient against 'all-hazard' potential threats." Although applicable to all CIP domains, in this case study, we discuss the concepts of *resilience* and *resilience innovation* in the context of electricity-related infrastructure for ensuring public safety and limiting supply disruption. For example, power grid outages occur for a number of reasons, including weather, aging infrastructure, and vandalism. Power outages and the required repairs to the power grid in the United States have been estimated by America's Electrical Cooperatives to be at least $177 billion in this coming decade because of the aging infrastructure (America's Electric Cooperatives 2017). The preventative maintenance and repairs prior to physical failure are essential to the reliability of the North American grid, and similar situations exist in other developed or developing nations.

Improving the resiliency of the power grid is critical given that the majority of other critical infrastructures sectors rely on consistent and reliable electricity. Notably, transportation networks are threatened by energy disruptions, the finance and banking sectors rely on electricity for operations, and telecommunication systems rely on continuous power. Food- and water-related resources are reliant on the energy infrastructure for their supply chains. Prolonged disruption to the electrical supply or distribution network can potentially result in significant economic losses. Preventative maintenance is imperative to ensuring uninterrupted supply.

Prevalence of Video Capture: Growing Opportunity for CIP

The supply of video in society is growing exponentially, from fixed surveillance cameras to wearable body cameras, cellphone cameras, and vehicular dashboard cameras. Unmanned aerial vehicles (UAVs) can provide another source of capturing video data, providing a previously untapped aerial perspective of the environment. Furthermore, with the impending availability of driverless cars and other autonomous and semiautonomous vehicles, a fleet of highly instrumented vehicles may soon become more common during deployment. Each vehicle continuously collects unprecedented amounts of sensor data. Although intended for the vehicle's functional purposes, such as navigation, this footage also has potential uses in CIP resilience analytics. Onboard sensors, such as cameras, have historically been used to monitor infrastructures (UAV America n.d.) and disasters (Allison et al. 2016). Machine learning algorithms and deep learning models can be leveraged in combination with these large data sources to monitor critical energy infrastructures to potentially detect failures before or to respond after they occur through actionable intelligence (Fig. 1).



(a) (b)

Fig. 1. Proposed surveillance system for monitoring energy sector-related infrastructures: (a) imaged scene could be used to detect power-related infrastructure and potentially detect and flag infrastructure at risk of failing; and (b) UAV-mounted cameras could be used for monitoring and assessment of rural or remote infrastructures.

Machine Vision for Image and Video Analysis

Machine vision is the combination of machine learning, in which computer algorithms automatically learn patterns from example data, and image analysis, in which objects of interest are automatically segmented and identified from video image data. Example applications include algorithms to segment regions of interest in medical computer-aided diagnostic systems (Sathish et al. 2016), the detection of oil spills from satellite imagery (Kubat et al. 1998), and the characterization of urban areas (Taubenböck et al. 2010). This technology has the potential to automate threat assessments in the context of securing critical infrastructures. The pertinent concepts and relevance to critical energy infrastructure systems are introduced here; additional technical details of machine vision are discussed thereafter.

Machine vision image data are generally sourced from two contexts: controlled scenes and unconstrained scenes. Controlled scene analyses are akin to a set of measurements taken in a lab or images taken within the constraints of a passport photo, in which figure posture, face size, and lighting are explicitly defined. In this situation, the context and type of image is given; therefore, systems can be designed according to the prescribed conditions. Unconstrained video, such as video from a surveillance camera or from a UAV, is substantially more difficult to analyze given the unknowns. For example, an engineer developing a system to identify individuals within a secure area has very little prior knowledge and may not know *a priori* how people will appear within these images. Furthermore, each person is of varying height, is likely wearing different clothes, appears in different positions and poses relative to the camera, may be wearing concealing items (e.g., sunglasses, hats, and scarves), or may be captured under differing lighting conditions throughout the day and night. Therefore, development requires an adaptive or intelligent system that handles a broad range of variation in images.

Machine vision has successfully been applied to a wide variety of applications, such as automated line queue estimation at stores from security video footage and autonomous vehicles fusing visible light and infrared imaging with other sensor data including radar and ultrasound. These systems are slowly emerging in the field of critical infrastructure protection (CIP). In particular, a subfield of machine vision—deep learning—may offer an increasing ability to monitor the environment for CIP tasks.

What is Deep Learning?

Deep learning is rapidly emerging as an effective way to create and train highly accurate machine vision systems, requiring only labeled images as input. Relevant patterns are learned automatically from the data. What differentiates deep learning from previous, less successful forms of machine vision is that both the decision function and the data representation are simultaneously learned and optimized from the training data. The word *deep* refers to the high

J. Infrastruct. Syst., 2019, 25(2): 05019003

number of layers (between raw image input and final classification output) typically used in such a network.

In a classic artificial neural network, an explicit preprocessing step exists in which features are extracted from raw measurements made from the image prior to inputting these features into the network. The feature extraction step traditionally requires human expertise through the process of feature engineering. However, in deep learning, the features are instead learned in parallel with the decision rules.

Deep learning algorithms are similar to the human brain in that they are adaptable. These algorithms become more accurate as more training data are provided to them. Deep learning models are able to simultaneously learn representation and decision rules from the data, similar to biological organisms from which they are inspired. For example, a trained individual such as a security officer is able to rapidly spot objects of interest in their field of view and might be responsible for simultaneously monitoring video feeds arriving from multiple cameras or sensor data feeds. For an individual to rapidly identify these objects of interest, they may be exposed to practice feeds to hone their perceptive ability, and the officer's accuracy improves with time and experience. Deep learning mimics the human ability of learning characteristics, patterns, trends, and other factors and tends to improve as additional training data are provided to the framework.

Intersection of CIP and Machine Vision

Critical infrastructures are relied on for essential services, and any disruption has significant economic, safety, and security impacts. Such an infrastructure represents a significant investment in modern society given the need for uninterrupted supply. A clear need exists to ensure the resilience of a critical energy infrastructure and to leverage technological advance promises to save costs and develop increasingly robust systems. This paper discusses how machine vision and deep learning can be used in conjunction with other sensing technology to improve resiliency.

The detection of objects of interest and the prediction of failure points are important in preventative maintenance. An object of interest is defined as any anomaly that may pose a threat to the infrastructure. Examples of threats to roadside energy infrastructures include utility poles with too much lean, fallen trees, vegetation overgrowth, and foreign objects that are not part of the as-designed infrastructure. Preventative maintenance always requires regular monitoring using sensors, whether hard-wired or video recorded from a UAV (UAV America n.d.) or ground-based vehicle (Redzuwan et al. 2013). These monitoring systems lead to considerable cost savings and enable the ability to monitor the infrastructure at finer-grained temporal resolutions. However, many such systems still rely on the time-consuming human analysis of the collected video and on a human operator to pilot the UAV.

A further enhancement is to reduce the human work of reviewing the captured video to identify potential issues within an energy corridor. Whereas acquiring a video using UAVs flying along an energy corridor significantly reduce costs, the most time-consuming and error-prone task of identifying objects of interest remains to the human: recorded video must be manually reviewed by trained experts to identify possible threats. This factor represents a significant cost and a bottleneck to throughput, whether speed or coverage. In such situations, machine vision models can form the basis of powerful video data analysis systems. Some research groups are investigating the use of UAVs to monitor the electrical infrastructure, and preliminary findings show promise despite being limited to *post hoc* analysis (Dutta et al. 2015; Varghese et al. 2017). These models can be made autonomous in that they aim to identify all threats within an image or video, or semiautonomous in that machine vision highlights areas of interest within a larger human-in-the-loop system, and the final determination is left to a human expert.

Machine vision algorithms can also be trained to assess the state of repair of the physical infrastructure. Symptoms such as rusting, cracks, deformations, and other visible signs of wear can be automatically identified to allow energy infrastructure owners/operators to increase resiliency by prioritizing resources. Machine vision successes in the field of civil engineering have included bridge inspection systems (Oh et al. 2009), vibrational monitoring applications (Caetano et al. 2011), and related concrete detection systems (Zhu and Brilakis 2010), including recent successes in applying deep learning (Cha et al. 2017). Although many applications can easily leverage image or video footage, machine vision can further decrease the cost of analyzing such footage. This system leads to lower costs and increased resiliency.

Structure of the Paper

The structure of this paper is outlined as follows. We first detail the technical background of deep learning to create and deploy machine vision systems, followed by a discussion of a number of current applications of deep learning and machine vision in critical infrastructure protection and threat identification. Thereafter, we present detailed results of a proof-of-concept study using deep learning to identify the power infrastructure within images typical of those captured by an instrumented ground vehicle. Finally, conclusions and recommendations on the use of deep learning for critical infrastructure protection are summarized.

Deep Learning for Machine Vision

Deep learning is a subfield of machine learning inspired by the functioning of the human brain and biological neurons. The novelty of deep learning is its departure from the manual development of features in favor of automatic detection, known as representation learning. This ability to automatically learn features from any type of digital information (e.g., text, images, video, and audio) without requiring domain experts to handcraft these features (i.e., feature engineering) indicates that deep learning can be broadly applied to many domains. Specifically, for computer vision, deep learning can form systems for which the primary input data are digital images and video, whereby a network effectively extracts high dimensional characteristics from images to learn about components of interest within an annotated image. The ultimate goal is to subsequently identify these elements in unlabeled images. Two machine vision tasks are possible: image classification, through which the entire image is classified as one or more labels (e.g., whether a given image is classified as a truck) or image segmentation, through which subregions of the input region are separated and labeled (e.g., a truck appears in the bottom-right corner of the image).

The first computer vision models were inspired from the study of the visual cortex (the part of the brain that processes and analyzes light in organic vision) of animals to determine how the brain interprets the visual stimuli from the environment (Hubel and Wiesel 1962). The last few decades have seen steady improvements in computer vision models. Since 2010, the arrival of deep learning networks has brought remarkable progress in the development of highly accurate and robust computer vision systems. A broad range of benchmark data sets, state-of-the-art models, and software tools exist to develop these frameworks. This section reviews the literature on the development of computer vision systems to motivate their use in critical infrastructure protection applications.

Benchmark Data Sets

The field of computing requires standardized methods to evaluate how one program performs relative to all others. In the case of computer vision, the best way to evaluate the success of models is to apply and compare all methods to a common set of benchmark images. The MNIST data set has been widely used because it consists of 70,000 images of handwritten digits in a convenient format for various machine learning models (LeCun et al. 1998). Subsequently, Krizhevsky (2009) developed two data sets of small (32×32 pixel) labeled images, CIFAR-10 and CIFAR-100, to test image classification systems for ten and 100 classes, respectively. More recently, the ImageNet database with more than 14 million images has been widely used, and each image is labeled with one of 21,841 terms from the WordNet hierarchy (Krizhevsky et al. 2012).

Types of Neural Networks

A broad variety of neural network architectures and training approaches exist, each having strengths in specific problem domains. The simplest form of a neural network is the feedforward Artificial Neural Network (ANN). In ANN, information flows strictly forward from input nodes through one or more layers of hidden nodes to the output nodes. Recently, convolutional neural networks (CNN) have been used with significant success in a number of image classification and segmentation tasks. A CNN consists of a feedforward ANN; however, instead of each layer having the same behavior, the network is made up of several alternating convolutional and pooling layers, followed by the final fully connected layer. More specifically, the structure begins with a convolutional layer in which spatial filters are convolved with different overlapping subregions of the input image. These filters are learned during training. Each convolutional layer is followed by a pooling layer that spatially aggregates the outputs of the convolutional layers using a form of nonlinear downsampling. The pooling layer provides several benefits including reducing the number of free parameters in the model and increasing the model's invariance to the translation of features within the input image. A series of these convolutions and pooling patterns is repeated before fully connecting the outputs from the last pooling layer to a fully connected layer. A final loss layer combines the outputs of the fully connected layer to arrive at the class labels of interest. Fig. 2 depicts the process of deep learning with an example of an abstract network to illustrate how features at multiple hidden layers are represented with examples from the authors' proof-of-concept network.

CNNs have been shown to be highly effective for the machine vision task. Through the convolutional layers, key features pertaining to the object of interest are extracted from the raw image, such as edges, corners, and end-points (Li et al. 2016). The complexity and/or scale of these features can grow in subsequent convolutional layers. Although image distortions, noise, or lighting conditions of the image can significantly impede the image recognition process, deep neural networks such as CNNs have shown to be highly robust to image quality. This ability partially explains the widespread use of CNN within state-of-the-art computer vision models (Dodge and Karam 2016).

State-of-the-Art Deep Learning Models

One of the first CNNs to achieve wide popularity was AlexNet, originally developed by researchers from the University of Toronto for the ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2010. AlexNet has been used by various other researchers for different image classification problems, motivating the exploration of even deeper networks. In fact, during ILSVRC 2014, a Google research team implemented a deeper network, GoogLeNet, which improved the results of the image classification task. The Visual Geometry Group (VGG) at Oxford University has explored how the depth of a CNN affects the resulting accuracy in large-scale image recognition tasks (Simonyan and Zisserman 2014). This group subsequently improved on GoogLeNet's performance using a more complex network configuration and shallower networks (Szegedy et al. 2015). Subsequently, the Google research team produced a deeper CNN fully trained on the ImageNet data set, with the code name Inception (Szegedy et al. 2015). Other groups have reused the Inception CNN for a wide range of image classification tasks, and it is used in this study. When retaining the initial convolution and pooling layers that extract features from the input image, one can retrain only the final fully connected classification layer (i.e., softmax) to suit the particular classification

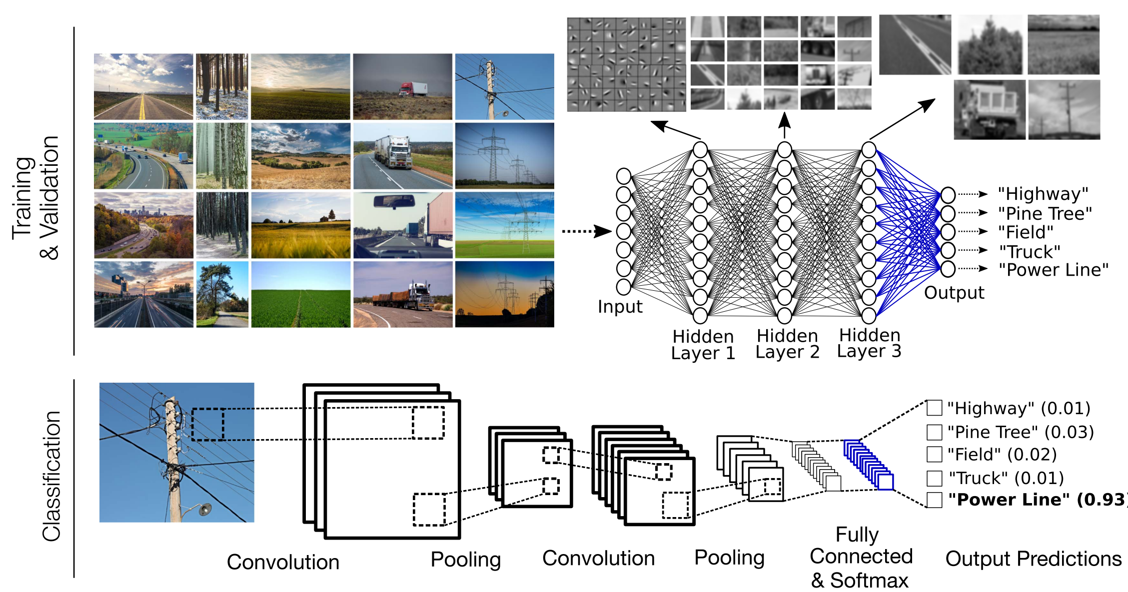


Fig. 2. Conceptual overview of convolutional neural network for CIP. (Images courtesy of Pixabay.)

problem at hand. Using this process, referred to as Transfer Learning, retraining the model and achieving significant savings in computational resources and training time is possible relative to completely retraining each layer (TensorFlow n.d.).

Hardware Implementation of Deep Networks

A primary concern with the implementation of deep neural networks is their computational requirements during training and deployment. Such networks may be characterized by thousands of individual parameters that must be learned or optimized from the training data; hence, significant computational requirements exist during training. However, once the network is trained and all of the parameters are optimized, the network can often be applied to very quickly analyze a new image. Computer vision systems require the manipulation of images (digitally represented as multidimensional matrices) and must perform large numbers of computations with these elements; such a task is ideally suited to the architecture of graphics processing units (GPUs) (Raina et al. 2009). Computer workstations or servers can now contain multiple GPUs; however, embedded systems seeking to use computer vision do not have the same capacity and flexibility to incorporate GPUs. For the purpose of deploying deep learning models for CIP applications, two fundamental approaches can be considered: (1) image data can be uploaded to the "cloud," where high-performance computational resources can be used to analyze the image data and actionable information returned; or (2) real-time, on-site deployment of computer vision models (on the "edge"), known as "Embedded Vision." These two approaches are discussed in the following sections.

In the Cloud: Offline Image Analysis

Machine vision models are currently most often deployed on large computation clusters or in the cloud for after-the-fact offline analysis. Traditional surveillance systems transmit images to a centralized location for analysis, requiring an expert human operator to analyze one or more video streams. This labor-intensive task is prone to high error rates given fatigue and a variety of human factors. However, in the case of computer vision systems, with sufficient processing power, any number of data streams can be analyzed simultaneously and in real-time and yet maintaining consistent performance. With this paradigm, all resources are allocated to a centralized location, information remains synchronized, and hardware can easily be maintained and monitored. Additionally, the centralization of large amounts of data could enable the computer vision model to progressively improve its performance.

For real-time processing, the images must be transmitted to a central server in real-time with minimal latency; therefore, the chosen communication channel must support the appropriate bandwidth level. Transmitting video data is potentially costly and may limit the applications and range of deployment of the sensor system (e.g., UAV) given the limitations inherent to the networking technology.

On the Edge: Embedded Vision Hardware Requirements

In mobile computer vision applications, such as UAVs, the onboard computing hardware is limited in size, weight, cost, and power. General central processing units (CPUs) are not ideally suited for machine vision tasks (Raina et al. 2009). Conversely, GPUs are well-suited to the image processing task but tend to have very high power requirements, making them ill-suited for smaller mobile applications (Wang and Chen 2012). Given these limitations, the trend for embedded vision systems has been toward creating dedicated embedded vision processors, wherein an application specific integrated circuit (ASIC) chip is developed purely for the purposes of rapid application of CNNs to input video in real time. As the field of machine vision grows, companies have begun to propose hardware improvements, and this sector is predicted to reach $300 billion (P. Paulin, "Applying Deep Learning Vision Technology to Low-Cost, Low-Power Embedded Systems: An Industrial Perspective," unpublished report). The development of dedicated embedded vision processors has been proposed, with the promise of lower power consumption, a smaller physical footprint, higher performance for the image analysis task, and the ability to incorporate a built-in, dedicated convolutional neural network engine (Conti et al. 2016; Park et al. 2016).

Such onboard systems will save on data transmission because only identified threats need to be communicated, even though doing so comes at the expense of increased onboard computational load. This situation represents both an increase in the cost of computer systems and higher power requirements. Lastly, the decentralization of data may affect future data-driven model improvements and the distributed update of models to embedded systems. Ultimately, operational requirements will determine the best-suited option.

Current Applications and Potential of Deep Learning in CIP

Significant ongoing research is occurring on the intersection of critical infrastructure protection and machine learning. Deep learning and machine vision are being developed for many CIP applications. Examples include roadside video content analytics (Verma et al. 2017), resilience to weather-related outages (Sun et al. 2016), tower detection (Liu et al. 2016), determination of structural integrity (Lee et al. 2018), food inspection (Marvin et al. 2017), and environmental safety (Mohanty et al. 2016). In the following sections, this paper discusses the current application of deep learning in several areas of CIP and reviews applications already employing machine vision and that have the potential to transition to deep learning approaches. A common theme throughout these sectors is that further developments in machine vision systems for CIP include improvements in video acquisition (e.g., using multiple overlapping cameras, autonomous or semiautonomous vehicles, or mobile UAVs) and video analysis, through which deep learning is expected to provide significant benefits.

Transportation

Thornton et al. (2009) examined current video analytics technologies with respect to CIP and, specifically, for transportation. They identify five categories of functionality, including low-level activity detection, high-level behavior detection, discrimination, tracking, and content retrieval, and evaluated existing technologies using a variety of video footage with staged events and behaviors representing threats to critical infrastructures. Their study highlights the difficulty of identifying threats in the presence of complex and time-varying backgrounds, the continued difficulty in automatically differentiating between humans and objects, and the complexity of tracking individuals across multiple camera fields of view. The use of video content analytics (VCA) in public transportation systems is common; however, VCA is effective only when the number of false positives (alarms needlessly triggered) is lower than the accepted threshold when still detecting events of interest. The limitations of fixed-point surveillance in rail transit systems have motivated the adoption of additional on-board cameras that improved performance over commercial VCA architectures across all test conditions (Casola et al. 2013). Several groups have recently demonstrated fast CNN models capable of interpreting complex

J. Infrastruct. Syst., 2019, 25(2): 05019003

roadside and rail-sourced scenes, indicating their robustness in environmental sensing tasks (Cai et al. 2016; Gibert et al. 2017; Yu et al. 2017). In addition, on-board camera surveillance drones have been proposed as a viable technology to reduce the cost and increase the granularity of the surveillance of rail transit infrastructures. This emerging technology offers smart-sensing functionalities that would be useful in railway safety operations, prognostics, forensic analysis, and general monitoring tasks (Flammini et al. 2016).

Electricity Distribution

In electrical utility infrastructures, resilience is being appraised with respect to physical failures, whether mechanical, vegetation, or weather-related outages. The Electric Power Research Institute weather incident management project seeks to better understand the impact of fallen vegetation on the power infrastructure (Zahodiakin 2016). Researchers have toppled trees across power lines and measured the impact forces that must be absorbed by poles and high-tension transmission lines. Forces exceeding the material strength of the pole result in material failures that lead to downed wires and blackouts.

In 2014, Sampedro et al. (2014) developed a supervised machine vision approach to electric tower detection and classification for power line inspections. Machine vision analysis of UAV-captured aerial images for the detection of power transmission lines has been investigated by a research team in Bangalore, India, and they achieved upwards of 99% accuracy (Dutta et al. 2015). Given the importance of inspecting power infrastructures and the associate expense, cost-saving innovations have been developed such as transmission line inspection robots (Wu et al. 2008, 2010). Although on-wire robots offer an unprecedented perspective, UAVs may be more effective because they can image the infrastructure from multiple viewpoints (Fig. 1). Recently, Varghese et al. (2017) reported preliminary findings in Bangalore, India of UAV-sourced imagery for damage detection of power-related infrastructures using a pre-trained CNN with promising results over a modest data set. These efforts corroborate this work and signal the emerging adoption of deep learning in different CIP sectors.

Critical Infrastructure Applications Using Machine Vision

Machine vision analysis has already been adopted to monitor various aspects of critical infrastructures, and these sectors have the potential to leverage deep learning systems. Machine vision is used in the structural assessment of concrete structures (Caetano et al. 2011; Lee et al. 2018; Oh et al. 2009; Zhu and Brilakis 2010); post-earthquake inspections to determine structural integrity and automated analysis are used to augment human qualitative observations (German et al. 2013); food security applications monitor goods such as poultry and apples for the detection of disease, defects, and contamination (Chen et al. 2002); and Blasco et al. (2002) proposed a robotic platform to identify and destroy weeds in agricultural settings.

Many existing and potential applications exist for machine vision in CIP. Deep learning has the potential to replace rule-based learning with complex automated reasoning. As the dominant emerging approach to creating machine vision systems, deep learning is expected to play an increasingly important role in CIP. To compliment the recent success of Varghese et al. (2017), who examined UAV-based imagery, the development of a proof-of-concept monitoring framework for vehicle-mounted surveillance of power infrastructures is presented in this paper.

Case Study: Deep Learning for Analysis of Vehicle-Mounted Video

With the emergence of semiautonomous and autonomous vehicles, it can be expected that the number of vehicles equipped with video acquisition hardware will continually increase. Even a nonautonomous vehicle equipped with a dashboard camera can capture large amounts of video data on the surrounding environments in which it navigates. Concurrently, with improvements in embedded vision system hardware, deep learning in mobile applications will become increasingly prevalent. This situation will create considerable opportunities for the implementation of automated monitoring technology for the purposes of CIP. With vehicles regularly driving through urban and rural areas, passive daily monitoring of infrastructures can potentially be achieved. Even remote areas with lower traffic volumes are regularly plied by commercial transport vehicles. Such vehicles provide the opportunity to function as passive monitors, relaying meaningful information on the status of road-related and nearby infrastructures. The work of the MIT Senseable City Lab exemplifies this and future potential through its City Scanner and Minimum Fleet projects (Anjomshoaa et al. 2018; Vazifeh et al. 2018).

Given steadily improving autonomous vehicle sensor and data processing hardware, a commensurate increase in the quality and quantity of video data available for object/event recognition systems is occurring that could be used in the automatic surveillance of environments. This example of resilience innovation could shift the CIP paradigm from reactionary toward preventative maintenance. Therefore, any infrastructure observable from a vehicle-mounted camera is a candidate for the development of a computer vision monitoring system (Table 1).

For critical infrastructures in less accessible regions, land access to monitor the structural integrity can be costly and resource intensive. If dashboard camera-sourced video was passively captured, a significant degree of monitoring could be achieved as other routine tasks and patrols are being fulfilled. Machine vision and real-time video capture can raise alarms to human operators when an infrastructure may have been compromised.

Whether vehicle- or drone-mounted, this augmentation of classical surveillance systems would result in cost reductions in the maintenance of critical infrastructures. A computer vision model may be able to detect more subtle changes in the environment that might escape the notice of human observers. Slow processes such as aging infrastructures often escape human perception because they occur on longer time scales. However, a computer vision model can determine when aging infrastructures reaches a

Table 1. Example applications of deep learning and vehicle-mounted cameras to CIP

Threat to critical infrastructure	Surveillance method	Environment
Compromised power infrastructure	Vehicular	Rural, remote
Graffiti	Vehicular	Urban
Damaged property	Vehicular, aerial	Urban, rural, remote
Blocked storm drain	Vehicular	Urban
Compromised/failing oil pipeline	Aerial	Remote
Rail debris or compromised structure	Train	Urban, rural, remote
Transportation truck clearance collision	Vehicular	Urban, rural, remote
Compromised road structural integrity	Vehicular	Urban, rural, remote

particular level and can even predict the time of failure and notify human operators.

This paper specifically examines the example of electricity corridors that are parallel to roadways. Maintenance of power infrastructures is costly and damaged or failed power infrastructures can result in considerable inconvenience and economic loss to numerous stakeholders, namely, those living and working within the vicinity of the failed infrastructure. This paper proposes a useful solution for the automatic detection of vegetation as a threat to critical power infrastructures. A study conducted on data collected from a U.S. electric utility company, Duke Power, estimated that trimming trees yearly decreases the number of power outages by a factor of 0.9 (Guikema et al. 2006). This work focuses on trees and branches as a particular type of debris that might compromise the functional integrity of the power infrastructure. However, the methods demonstrated herein can be broadly applied to other threats to critical infrastructures (Table 1).

Data Collection, Image Analysis, and Transfer Learning

This paper defines power-related infrastructures as power poles, power lines, power towers, and accessory power hardware such as power pole-mounted transformers and guy-wire. Considering that much of the power infrastructure is deployed along road or service road corridors, this paper proposes that vehicle-mounted cameras (e.g., dashboard cameras) could be used to automatically acquire images of such an infrastructure. Subsequent image analysis using deep learning and machine vision can identify potential threats to this infrastructure, such as branches nearing power corridors, fallen trees, or other debris hanging on power lines.

The demonstrated power of CNNs in image classification and segmentation tasks motivates their use to solve the detection task in localizing power-related infrastructures in the acquired images. Consider that a simple tree trunk can easily be mistaken as a power pole during the recognition process when aiming to detect a power line. Even expert humans may sometimes have difficulty with such challenging classifications. To overcome this issue in machine vision, using a deep network with multiple layers becomes necessary to differentiate these otherwise similar objects and correctly identify power-related infrastructures. The TensorFlow version 1.0 deep learning framework, a software toolkit released by Google in late 2015, was used in this work and has been widely adopted as a primary tool in the development of CNNs (Abadi et al. 2016). All models were developed and evaluated using SHARCNET high performance GPU computing infrastructure (sharcnet.ca). An IBM S822LC server with dual power8+ chips, 10 cores per socket, 8 SMT (Simultaneous MultiThreading) pipelines per core, and 4 NVIDIA Pascal P100 GPUs connected with NVlinks was used.

This paper develops a proof-of-concept model to establish that deep learning is indeed applicable to the task of surveillance of power-related infrastructures through images from vehicle-mounted cameras. The notion of Transfer Learning, as previously introduced, implies that a pretrained deep network can reweight its final two layers to solve a new classification problem in a different image domain. Meanwhile, the initial layers of the network, which extract the image representation used for classification, are reused without requiring retraining. This approach is particularly useful for cases in which the data set is limited in size, as is the case presented in this paper. Here, Transfer Learning was applied to the Inception (v3) CNN.

Five classes were selected that corresponded to the most prevalent elements within this image data set to classify among highways, pine trees, fields, trucks, and power-related infrastructures. To train the model, the authors acquired images corresponding to these five classes from the ImageNet data set. Approximately 1,200 images from each class were initially selected. Images captured at night were manually removed because the current implementation assumed that video data are captured during daylight hours. Images found to be irrelevant to the class were also removed after a cursory visual inspection. The final image composition of the data set used in the Transfer Learning task was comprised of 4,053 images: 723 Highway, 876 Pine Tree, 873 Field, 564 Truck, and 1,017 Power Line. Sample images for each class are depicted in Fig. 3.

High-Accuracy Results for Five Class Classification Task

The experimental design of this study considered three data set configurations to evaluate the impact of differently sized training,

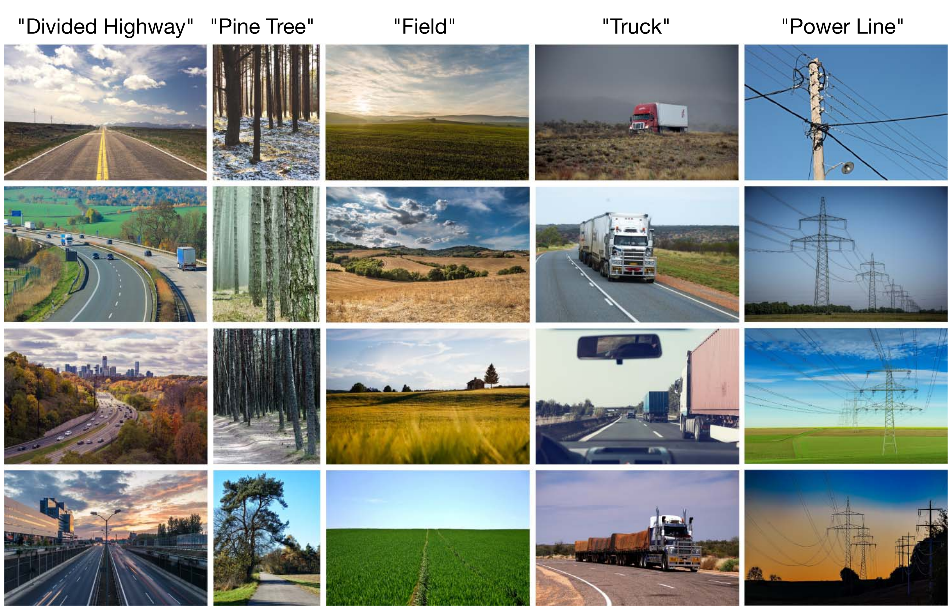


Fig. 3. Sample images of five classes from ImageNet. (Images courtesy of Pixabay.)

Table 2. Performance metrics of inception v3 model retrained using transfer learning on set of images across five classes: highway, pine tree, field, truck, and power line

Data set configuration[a]	Epochs	Train accuracy (%)	Cross entropy	Validation accuracy[b] (%)	Final test accuracy (%)	Final test set size
80:10:10	4,000	99.0	0.089	94.0	94.7	377
	5,000	99.0	0.062	96.0	94.4	
	10,000	100	0.046	93.0	94.2	
70:15:15	4,000	98.0	0.059	95.0	95.0	619
	5,000	97.0	0.093	93.0	95.2	
	10,000	**100**	**0.036**	**96.0**	**95.5**	
60:20:20	4,000	99.0	0.070	95.0	94.8	826
	5,000	98.0	0.074	95.0	95.0	
	10,000	100	0.043	97.0	95.0	

Note: Bold values indicate the final selected model.
[a]Format—train:validation:test.
[b]N = 100.

validation, and test data sets on model accuracy. For example, the first configuration used from each class 80% of the images in training, 10% in validation, and 10% in testing. We simplified the notation of these configurations to train:validation:test for ease of reference. These three configurations were: 80:10:10, 70:15:15, and 60:20:20. We further examined the influence of training time by considering three values for the number of epochs: 4,000, 5,000, and 10,000. Together, we compared the classification performance over nine models to highlight the highest performing configuration as our final model.

Each data set configuration was comprised of an independent random shuffling of images into one of three subsets. Deep network training was run for 4,000 training epochs, increased to 5,000, and then to 10,000, with an instance of the model saved at each time point. During each epoch, ten images were randomly selected from the training set and used to update the final layer of the network using back-propagation. A validation set was used after each epoch to ensure that the network was not overfitting to the training data. For the three data set configurations, approximately 10%, 15%, and 20% of the total data were respectively withheld (377, 619, and 826 images; stratified sample) as a holdout test set to assess the final performance of the retrained classifier. Performance across the nine models is summarized in Table 2. Although the performance is largely consistent among all models, the classifier trained and tested on the 70:15:15 split following 10,000 epochs led to the highest final test accuracy, which is emboldened in Table 2.

When comparing the classification accuracy of these models, we observed an unsubstantial decline in the final test accuracy with increased training time of the models trained on the 80:10:10 split. However, given increased sizes in the validation and test sets of the two other configurations, we observed modest overall improvement in the final test accuracy with additional training time. Furthermore, the fact that classification accuracy is maintained from the validation set to the holdout set across all models gives us confidence that these models are not overfit to the data.

To further explore the system performance, a full confusion table over the 619 images of the holdout test set of our selected model is provided in Table 3. Performance was strongest on the Truck and Power Line classes, and the weakest performance was on the Field class. Considering that many of the images from the Field class also contained portions of roadways, vehicles, animals, and trees, that this heterogeneous class was difficult to classify is not surprising. Such complex realistic images suggest that a multilabel classification strategy may be appropriate, in which each image can actually receive multiple correct labels. The corresponding confusion tables for the remaining eight models are available in Supplemental Figs. S1–S8.

Interestingly, the results obtained from 4,000 to 10,000 epochs of the 80:10:10 configuration indicate a clear decline in cross entropy and a modest rise in training accuracy. These results alone might mislead someone to conclude that increasing the number of epochs is beneficial for the general accuracy of the model. However, when considering the final test accuracy, we note that a larger number of epochs decreases this value, revealing significant misclassifications. As the model continued to be trained, differentiating the Field class from the Power Line class became increasingly challenging after 5,000 epochs (five misclassifications). After 10,000 epochs, the Field class was further confused by both the Power Line and Pine Tree classes (eight misclassifications). This misclassification could be explained by the similar composition of the scenes represented by the classes Field, Power Line and Pine Tree as opposed to Truck, which is easily differentiated with the presence of a truck in the image.

Table 3. Full confusion matrix after 10,000 epochs over the final hold-out test set (619 images)

	Predicted classes					
Actual classes	Pine	Highway	Truck	Field	Power line	Total (%)
Pine tree	**115**	0	0	3	0	**97.46**
Highway	0	**109**	3	1	4	**93.16**
Truck	0	0	**78**	0	0	**100.0**
Field	3	0	2	**147**	1	**96.08**
Power line	4	2	2	3	**142**	**92.81**
Total (%)	**94.26**	**98.20**	**91.76**	**95.45**	**95.30**	**95.48**

Note: Bold values indicate results specific to each class.

In contrast to the first configuration, the two other data set splits that attributed a larger proportion of images to the test and validation sets each demonstrated a consistent overall increase in performance and a modest increase in the final test accuracy with increased training time, although the 60:20:20 models ultimately plateaued. Examination of the misclassifications across all models revealed many of the same images, indicative of the complex compositions of these images. The model generated from the 70:15:15 configuration following 10,000 iterations led to the highest accuracy on our holdout image set with an acceptable number of misclassifications; therefore, this model defines the hyperparameters to be used in future work.

Discussion and Future Work

In this paper, the adoption of deep learning was proposed throughout the sectors of CIP. This proposal represents a shift from reactionary maintenance in the case of power failures and toward preventative maintenance to create increasingly resilient power infrastructure systems. Overall, the results have demonstrated that deep learning can effectively identify power-related critical infrastructures within images representative of vehicle-mounted cameras. The following sections describe an incremental approach to incorporating deep learning models to passively monitor critical infrastructures.

From Image Classification to Image Segmentation

The success of the proof-of-concept image classification model indicates that the CNN can distinguish among five images classes representative of images that are expected to be collected from vehicle-mounted cameras in rural settings. The next stage is to move from image classification, which assigns an entire image a single classification, to image segmentation, which delineates and identifies discrete objects within a complex image. An example of an idealized image segmentation can be observed in Fig. 4; the output image from an image segmentation task highlights those pixels in the image corresponding to each of the classes that it detects within the image and each respectively bounded: Highway, Tree, Field, Vehicle, and Power Infrastructure. The segmentation system can again use the Inception v3 model; however, the final classification layer (i.e., softmax) will be modified in such a way that various regions of an input image are classified to distinguish the sections of the image that correspond to each class label.

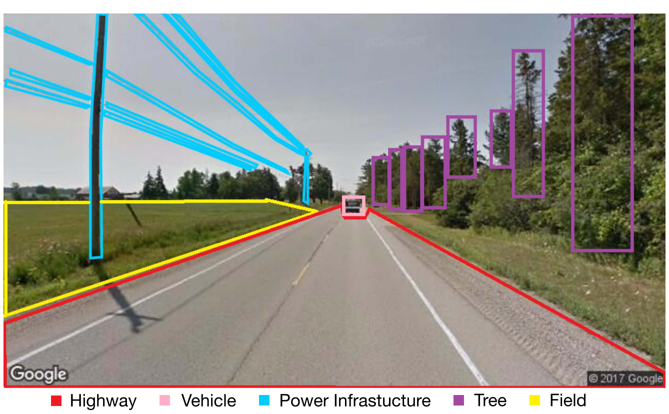


Fig. 4. Idealized image segmentation output across five classes. (Image © 2017 Google.)

This work will require a curated data set of manually segmented images. Considering the focus on critical electrical infrastructures deployed along roadways or accessible right-of-ways, the method requires images typical of such environments. Most roadways have been extensively photographed by Google for their Google Street View product. The images are freely available in limited quantities and the entire database is also available for purchase. Alternative open-source methods can be used to systematically sample these data (Dick et al. 2018b) and amass a database of a number of electricity corridors typical of a rural landscape as in Dick et al. (2018a). The manual gold-standard annotation of the elements within each image of the data set requires an annotation tool such as LabelMe (Russell et al. 2008) to generate logical masks corresponding to regions containing a certain class, as in Varghese et al. (2017). The use of Google Street View images and the LabelMe software could be leveraged to generate data sets useful in training and deploying an image segmentation model. Power infrastructures not adjacent to public roads could be collected along service corridors during routine patrols on company vehicles or by using UAVs, as explored in Varghese et al. (2017).

Fully-Trained CNN for Analysis of Sequential Frames from Video

The development of CNN for image classification and segmentation discussed to this point have leveraged Transfer Learning applied to an established CNN. This enables the rapid development of a computer vision model using a relatively small data set of images in a resource-efficient manner. However, the results generated by that model are subject to the training process of the network. In the deployment proposed in this work, images to be analyzed would be extracted from video and hence would represent temporally and spatially sequential frames. Since Inception v3 was not trained on such sequential data, it has no ability to interpret and leverage minute changes between sequential images. Instead, all images are treated independently and therefore temporal information is lost. To rectify this, development of a dedicated CNN for the explicit task of power-related infrastructure detection based on video streams from dashboard cameras is proposed. This approach will enable the capture of large amounts of image data to create a sizable data set on which to train the model. Other deep network architectures have been applied to sequential data, such as the use of Long Short-Term Memory models (Hochreiter and Schmidhuber 1997) for machine translation, image classification, and natural language processing. Architectures such as these should be explored for developing video-based frameworks as opposed to the strictly image-based systems discussed here.

CNN for Detecting Threats to Critical infrastructures

Ultimately, once models have been developed and validated for identifying objects of interest from images, such as power lines, poles, trees, and others, the next step is to then develop a system for identifying situations that pose a threat to critical energy infrastructure. A new training and evaluation data set of video-derived sequential images will be created. When generating this data set, a number of threatened infrastructure cases will be flagged as "positives" such that the network can be trained to detect them. Examples of curated images include those in which trees intersect the power lines, poles lean at critical angles, or physical damage is visibly present.

Building on the highly accurate proof-of-concept network presented in this paper, future work can focus on developing an image segmentation algorithm, leveraging these deep networks to identify and isolate power lines and infrastructures from the image. Possible applications include identifying fallen trees hanging on power lines to flag areas in which branches and foliage approach power line corridors, and to identify infrastructure that appears to be failing, such as a badly leaning pole or tower.

Toward Model Deployment

This preliminary work shows considerable promise in applying deep learning to the detection of power-related infrastructures; however, these experiments were designed with approximately balanced classes. In reality, we expect the number of "at risk" instances (positives) in our data sets to be very low relative to the number of "normal" instances (negatives), thereby exhibiting substantial and dynamic class imbalance. This situation represents an interesting paradox within this line of research: we seek to develop highly accurate models capable of detecting the very rare events of failing infrastructures to better maintain these systems, making these events rarer still and the challenge of detecting them all the greater.

Should the prevalence of failing infrastructure be on the order of one instance for every thousand, even models capable of achieving 99.9% accuracy would be useless in practice. To be useful, deployable models must be trained to optimize metrics appropriate to this level of class imbalance, such as maximizing the area under the prevalence-corrected precision-recall curve. Ultimately, such a solution can only be deployed if it exhibits performance metrics similar to or surpassing what is currently achievable by human operators.

Conclusions and Recommendations

Critical infrastructures are ubiquitously present in daily life and must be maintained and protected against threats. Energy infrastructures are intrinsically linked to other sectors of critical infrastructures and their supply disruption can have catastrophic cascading consequences. Deep learning and machine vision are promising new technologies that may form the basis for solutions to improve the resilience of critical energy infrastructures. This paper has summarized the current state-of-the-art in deep learning for creating machine vision models and has investigated the current applications of this technology to CIP. The proposed application of machine vision to analyze images acquired from vehicle-mounted cameras for critical energy infrastructure protection was discussed in detail. A proof-of-concept deep learning machine vision system was developed which can identify energy infrastructure with high accuracy, achieving 95.5% in a five-class classification task.

Moving forward, a number of issues are to be addressed before systems are deployed in an official infrastructure monitoring capacity. This paper has highlighted considerations that must be addressed to develop robust applications capable of unsupervised deployment, such as the requirement for fine-grained ground truth data, centralization of data, and iterative model improvement. Following the process of resilience innovation, advances in machine vision and deep learning will synergize with new data gathering mechanisms, such as instrumented autonomous vehicles or UAVs, to augment our capacity for monitoring threats to critical infrastructures.

Supplemental Data

Tables S1–S8 are available online in the ASCE Library (www.ascelibrary.org).

References

Abadi, M., et al. 2016. "TensorFlow: Large-scale machine learning on heterogeneous distributed systems." Preprint, submitted March 14, 2016. https://arxiv.org/abs/1603.04467.

Allison, R., J. Johnston, G. Craig, and S. Jennings. 2016. "Airborne optical and thermal remote sensing for wildfire detection and monitoring." *Sensors* 16 (8): 1310. https://doi.org/10.3390/s16081310.

America's Electric Cooperatives. 2017. "Can the infrastructure be fixed?" Accessed March 20, 2017. http://www.electric.coop/.

Anjomshoaa, A., F. Duarte, D. Rennings, T. Matarazzo, P. de Souza, and C. Ratti. 2018. "City scanner: Building and scheduling a mobile sensing platform for smart city services." *IEEE Internet Things J.* 5 (6): 4567–4579. https://doi.org/10.1109/JIOT.2018.2839058.

Blasco, J., N. Aleixos, J. M. Roger, G. Rabatel, and E. Moltó. 2002. "AE—Automation and emerging technologies: Robotic weed control using machine vision." *Biosyst. Eng.* 83 (2): 149–157. https://doi.org/10.1006/bioe.2002.0109.

Caetano, E., S. Silva, and J. Bateira. 2011. "A vision system for vibration monitoring of civil engineering structures." *Exp. Tech.* 35 (4): 74–82. https://doi.org/10.1111/j.1747-1567.2010.00653.x.

Cai, Y., X. Sun, H. Wang, L. Chen, and H. Jiang. 2016. "Night-time vehicle detection algorithm based on visual saliency and deep learning." *J. Sensors* 2016: 1–7. https://doi.org/10.1155/2016/8046529.

Casola, V., M. Esposito, F. Flammini, N. Mazzocca, and C. Pragliola. 2013. "Performance evaluation of video analytics for surveillance on-board trains." In *Proc., Int. Conf. on Advanced Concepts for Intelligent Vision Systems*, 414–425. Cham, Switzerland: Springer.

Cha, Y.-J., W. Choi, and O. Büyüköztürk. 2017. "Deep learning-based crack damage detection using convolutional neural networks." *Comput.-Aided Civ. Infrastruct. Eng.* 32 (5): 361–378. https://doi.org/10.1111/mice.12263.

Chen, Y.-R., K. Chao, and M. S. Kim. 2002. "Machine vision technology for agricultural applications." *Comput. Electron. Agric.* 36 (2–3): 173–191. https://doi.org/10.1016/S0168-1699(02)00100-X.

Community and Regional Resilience Institute. 2013. *Definitions of community resilience: An analysis*. Washington, DC: Community and Regional Resilience Institute.

Conti, F., D. Rossi, A. Pullini, I. Loi, and L. Benini. 2016. "PULP: A ultra-low power parallel accelerator for energy-efficient and flexible embedded vision." *J. Signal Process. Syst.* 84 (3): 339–354. https://doi.org/10.1007/s11265-015-1070-9.

Dick, K., F. Charih, and J. Green. 2018a. "High quality annotations of power infrastructure in Rural Ontario." In *Proc., Scholars Portal Dataverse*. Toronto: Scholars Portal Dataverse.

Dick, K., F. Charih, Y. Souley Dosso, L. Russell, and J. R. Green. 2018b. "Systematic street view sampling." In *Proc., 15th Conf. on Computer and Robot Vision*. Piscataway, NJ: IEEE.

Dodge, S., and L. Karam. 2016. "Understanding how image quality affects deep neural networks." In *Proc., 8th Int. Conf. on Quality of Multimedia Experience, QoMEX 2016*. Lisbon, Portugal: IEEE Signal Processing Society.

Dutta, T., H. Sharma, A. Vellaiappan, and P. Balamuralidhar. 2015. "Image analysis-based automatic detection of transmission towers using aerial imagery." In *Proc., Iberian Conf. on Pattern Recognition and Image Analysis*, 641–651. Cham, Switzerland: Springer.

Flammini, F., R. Naddei, C. Pragliola, and G. Smarra. 2016. "Towards automated drone surveillance in railways: State-of-the-art and future directions." In *Proc., Int. Conf. on Advanced Concepts for Intelligent Vision Systems*, 336–348. Lecce, Italy: Springer.

German, S., J.-S. Jeon, Z. Zhu, C. Bearman, I. Brilakis, R. DesRoches, and L. Lowes. 2013. "Machine vision-enhanced postearthquake inspection." *J. Comput. Civ. Eng.* 27 (6): 622–634. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000333.

Gibert, X., V. M. Patel, and R. Chellappa. 2017. "Deep multitask learning for railway track inspection." *IEEE Trans. Intell. Transp. Syst.* 18 (1): 153–164. https://doi.org/10.1109/TITS.2016.2568758.

Government of Canada. 2017. "Critical infrastructure." Accessed June 6, 2017. https://www.publicsafety.gc.ca/cnt/ntnl-scrt/crtcl-nfrstrctr/index-en.aspx.

Guikema, S. D., R. A. Davidson, and H. Liu. 2006. "Statistical models of the effects of tree trimming on power system outages." *IEEE Trans. Power Delivery* 21 (3): 1549–1557. https://doi.org/10.1109/TPWRD.2005.860238.

Hochreiter, S., and J. Schmidhuber. 1997. "Long short-term memory" *Neural Comput.* 9 (8): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hubel, D. H., and T. N. Wiesel. 1962. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." *J. Physiol.* 160 (1): 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837.

Krizhevsky, A. 2009. "CIFAR-10 and CIFAR-100 datasets." Accessed February 1, 2018. https://www.cs.toronto.edu/~kriz/cifar.html.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "ImageNet classification with deep convolutional neural networks." In *Proc., Advances in Neural Information Processing Systems 25*, 1097–1105. Red Hook, NY: Curran Associates, Inc.

Kubat, M., R. C. Holte, and S. Matwin. 1998. "Machine learning for the detection of oil spills in satellite radar images." *Mach. Learn.* 30 (2/3): 195–215. https://doi.org/10.1023/A:1007452223027.

LeCun, Y., C. Cortes, and C. Burges. 1998 "MNIST handwritten digit database." Accessed March 21, 2017. http://yann.lecun.com/exdb/mnist/.

Lee, S., J. Ha, M. Zokhirova, H. Moon, and J. Lee. 2018. "Background information of deep learning for structural engineering." *Arch. Comput. Methods Eng.* 25 (1): 121–129. https://doi.org/10.1007/s11831-017-9237-0.

Li, S., Y. Dou, X. Niu, Q. Lv, and Q. Wang. 2016. "A fast and memory saved GPU acceleration algorithm of convolutional neural networks for target detection." *Neurocomputing* 230: 48–59. https://doi.org/10.1016/j.neucom.2016.11.046.

Liu, Y., J. Yong, L. Liu, J. Zhao, and Z. Li. 2016. "The method of insulator recognition based on deep learning." In *Proc., 4th Int. Conf. on Applied Robotics for the Power Industry (CARPI)*, 1–5. Piscataway, NJ: IEEE.

Marvin, H. J. P., E. M. Janssen, Y. Bouzembrak, P. J. M. Hendriksen, and M. Staats. 2017. "Big data in food safety: An overview." *Crit. Rev. Food Sci. Nutr.* 57 (11): 2286–2295. https://doi.org/10.1080/10408398.2016.1257481.

Mohanty, S. P., D. P. Hughes, and M. Salathé. 2016. "Using deep learning for image-based plant disease detection." *Front. Plant Sci.* 7: 1419. https://doi.org/10.3389/fpls.2016.01419.

Oh, J.-K., G. Jang, S. Oh, J. H. Lee, B.-J. Yi, Y. S. Moon, J. S. Lee, and Y. Choi. 2009. "Bridge inspection robot system with machine vision." *Autom. Constr.* 18 (7): 929–941. https://doi.org/10.1016/j.autcon.2009.04.003.

Park, S., I. Hong, J. Park, and H.-J. Yoo. 2016. "An energy-efficient embedded deep neural network processor for high speed visual attention in mobile vision recognition SoC." *IEEE J. Solid-State Circuits* 51 (10): 2380–2388.

Public Safety Canada. 2015. "Enhancing critical infrastructure resiliency." Accessed June 6, 2017. https://www.publicsafety.gc.ca/cnt/ntnl-scrt/crtcl-nfrstrctr/nhncng-rslnc-en.aspx.

Raina, R., A. Madhavan, and A. Y. Ng. 2009. "Large-scale deep unsupervised learning using graphics processors." In *Proc., 26th Annual Int. Conf. on Machine Learning—ICML '09*, 1–8. New York: Association for Computing Machinery.

Redzuwan, R. M., N. M. Din, M. Z. Baharuddin, I. S. Mustafa, and R. C. Omar. 2013. "Transmission tower environment monitoring using UAV." In Vol. 16 of *Proc., IOP Conf. Series: Earth and Environmental Science*, 012105. Bristol, UK: IOP Publishing.

Russell, B. C., A. Torralba, K. P. Murphy, and W. T. Freeman. 2008. "LabelMe: A database and web-based tool for image annotation." *Int. J. Comput. Vision* 77 (1–3): 157–173. https://doi.org/10.1007/s11263-007-0090-8.

Sampedro, C., C. Martinez, A. Chauhan, and P. Campoy. 2014. "A supervised approach to electric tower detection and classification for power line inspection." In *Proc., Int. Joint Conf. on Neural Networks*, 1970–1977. Piscataway, NJ: IEEE.

Sathish, D., S. Kamath, K. V. Rajagopal, and K. Prasad. 2016. "Medical imaging techniques and computer aided diagnostic approaches for the detection of breast cancer with an emphasis on thermography: A review." *Int. J. Med. Eng. Inf.* 8 (3): 275.

Simonyan, K., and A. Zisserman. 2014. "Very deep convolutional networks for large-scale image recognition." Preprint, submitted September 4, 2014. https://arxiv.org/abs/1409.1556.

Sun, H., Z. Wang, J. Wang, Z. Huang, N. Carrington, and J. Liao. 2016. "Data-driven power outage detection by social sensors." *IEEE Trans. Smart Grid* 7 (5): 2516–2524. https://doi.org/10.1109/TSG.2016.2546181.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. 2015. "Going deeper with convolutions." In *Proc., IEEE Conf. on Computer Vision and Pattern Recognition*, 1–9. Piscataway, NJ: IEEE.

Taubenböck, H., T. Esch, M. Wurm, A. Roth, and S. Dech. 2010. "Object-based feature extraction using high spatial resolution satellite data of urban areas." *J. Spatial Sci.* 55 (1): 117–132. https://doi.org/10.1080/14498596.2010.487854.

TensorFlow. n.d. "How to retrain inception's final layer for new categories." Accessed March 22, 2017. https://www.tensorflow.org/tutorials/image_retraining.

Thornton, J., J. Baran-Gale, and A. Yahr. 2009. "An assessment of the video analytics technology gap for transportation facilities." In *Proc., 2009 IEEE Conf. on Technologies for Homeland Security*, 135–142. Piscataway, NJ: IEEE.

UAV America. n.d. "Electrical grid inspection." Accessed March 30, 2017. http://uavamerica.com/electrical-grid-inspection/.

Varghese, A., J. Gubbi, H. Sharma, and P. Balamuralidhar. 2017. "Power infrastructure monitoring and damage detection using drone captured images." In *Proc., 2017 Int. Joint Conf. on Neural Networks (IJCNN)*, 1681–1687. Piscataway, NJ: IEEE.

Vazifeh, M. M., P. Santi, G. Resta, S. H. Strogatz, and C. Ratti. 2018. "Addressing the minimum fleet problem in on-demand urban mobility." *Nature* 557 (7706): 534–538. https://doi.org/10.1038/s41586-018-0095-1.

Verma, B., L. Zhang, and D. Stockwell. 2017. "Deep learning techniques for roadside video data analysis." In *Roadside video data analysis*, 119–157. Singapore: Springer.

Victor, D. G., and K. Yanosek. 2017. "The next energy revolution: The promise and peril of high-tech innovation." *Foreign Aff.* 96: 124.

Wang, H., and Q. Chen. 2012. "An energy consumption model for GPU computing at instruction level." 4 (2): 192–200.

Wu, G., H. Xiao, X. Xiao, Z. Huang, and Y. Li. 2010. "Transmission line inspection robot and deicing robot: Key technologies, prototypes and applications." In *Proc., 1st Int. Conf. on Applied Robotics for the Power Industry, CARPI 2010*, 1–6. Piscataway, NJ: IEEE.

Wu, G., X. Xu, H. Xiao, J. Dai, X. Xiao, Z. Huang, and L. Ruan. 2008. "A novel self-navigated inspection robot along high-voltage power transmission line and its application." In *Proc., Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1145–1154. Berlin, Heidelberg: Springer.

Yu, S., Y. Wu, W. Li, Z. Song, and W. Zeng. 2017. "A model for fine-grained vehicle classification based on deep learning." *Neurocomputing* 257: 97–103. https://doi.org/10.1016/j.neucom.2016.09.116.

Zahodiakin, P. 2016. "Three-year research project looks at pole-top components, vegetation management, and more." Accessed March 20, 2017. http://eprijournal.com/making-distribution-grids-stronger-more-resilient/.

Zhu, Z., and I. Brilakis. 2010. "Machine vision-based concrete surface quality assessment." *J. Constr. Eng. Manage.* 136 (2): 210–218. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000126.