Context Integration for Reliable Anomaly Detection from Imagery Data for Supporting Civil Infrastructure Operation and Maintenance

by

Jiawei Chen

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2020 by the
Graduate Supervisory Committee:

Pingbo Tang, Chair
Steven Ayer
Yezhou Yang

ARIZONA STATE UNIVERSITY

August 2020

ABSTRACT

Imagery data has become important for civil infrastructure operation and maintenance because imagery data can capture detailed visual information with high frequencies. Computer vision can be useful for acquiring spatiotemporal details to support the timely maintenance of critical civil infrastructures that serve society. Some examples include: irrigation canals need to maintain the leaking sections to avoid water loss; project engineers need to identify the deviating parts of the workflow to have the project finished on time and within budget; detecting abnormal behaviors of air traffic controllers is necessary to reduce operational errors and avoid air traffic accidents. Identifying the outliers of the civil infrastructure can help engineers focus on targeted areas. However, large amounts of imagery data bring the difficulty of information overloading. Anomaly detection combined with contextual knowledge could help address such information overloading to support the operation and maintenance of civil infrastructures.

Some challenges make such identification of anomalies difficult. The first challenge is that diverse large civil infrastructures span among various geospatial environments so that previous algorithms cannot handle anomaly detection of civil infrastructures in different environments. The second challenge is that the crowded and rapidly changing workspaces can cause difficulties for the reliable detection of deviating parts of the workflow. The third challenge is that limited studies examined how to detect abnormal behaviors for diverse people in a real-time and non-intrusive manner. Using video and

relevant data sources (e.g., biometric and communication data) could be promising but still need a baseline of normal behaviors for outlier detection.

This dissertation presents an anomaly detection framework that uses contextual knowledge, contextual information, and contextual data for filtering visual information extracted by computer vision techniques (ADCV) to address the challenges described above. The framework categorizes the anomaly detection of civil infrastructures into two categories: with and without a baseline of normal events. The author uses three case studies to illustrate how the developed approaches can address ADCV challenges in different categories of anomaly detection. Detailed data collection and experiments validate the developed ADCV approaches.

DEDICATION

I dedicate this dissertation to my parents, Hong Chen, and Yingqiong Yang. They support me in higher education. I owe a lot to their love, support, and trust in me. I cannot accomplish this dissertation without their help.

ACKNOWLEDGMENTS

Upon the completion of this dissertation, I cannot imagine that I can finish it without the help and support from my committee, Dr. Pingbo Tang (Chair), Dr. Steven Ayer, and Dr. Yezhou Yang. I want to express my most profound appreciation for your continuous help and support, which makes me become a better researcher.

My most profound appreciation goes to my advisor Dr. Pingbo Tang. I appreciate him for giving me this precious opportunity of joining his research group and support me in the Ph.D. program. I sincerely thank him for his encouragement and for supporting my career goals. He spent lots of time guiding me and teaching me how to be a good researcher. He always shared lots of experience and knowledge that benefit me a lot. I am fortunate to have him as my advisor.

Also, I want to thank the members of my committee, Dr. Steven Ayer, and Dr. Yezhou Yang. They helped me a lot with the completion of my dissertation. Dr. Steven is always helpful in providing constructive suggestions for my thesis and presentation. Dr. Yang is an expert in the computer science domain and helped me learn a lot in computer vision. Moreover, all my committee spent much time reviewing my thesis and discussing my presentation slides. I appreciate their help and support.

I am also grateful for all my colleagues and friends at Arizona State University, especially Cheng Zhang, Vamsi, Zhe Sun, Yanyu Wang, Jinding Xing, and Ying Shi for their intellectual and emotional support.

Finally, words cannot express my appreciation to my family. My parents always believed in me and took pride in me. Their trust and support are the most important things that motivate me to overcome the difficulties in my research.

This dissertation presents the work supported by the Joint Research Program of Salt River Project (SRP), the U.S. Department of Energy (DOE), Nuclear Engineering University Program (NEUP) under Award No. DENE0008403 and NASA University Leadership Initiative program (Contract No. NNX17AJ86A, Project Officer: Dr. Anupa Bajwa, Principal Investigator: Dr. Yongming Liu). SRP, DOE, NASA's support is gratefully acknowledged. Any opinions, findings, conclusions, or recommendations expressed in this dissertation are those of the author and do not necessarily reflect the reviews of SRP, DOE, NASA, or Arizona State University.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

1 INTRODUCTION

Timely maintenance of critical civil infrastructures is essential to ensure adequate infrastructure operation and maintenance. These facilities are well past their design lives (ACSE, 2013). For instance, the waterways in the U.S. serve for moving goods across the country. There are 257 locks currently in use in the inland waterways, and 30 of these locks are from the 1800s. Another 92 are in service for over 60 years, while the design life of a typical lock is 50 years (ASCE, 2013). Timely maintenance of these aging infrastructures is, therefore, necessary. According to previous studies, imagery data has become one significant data source for civil infrastructure condition assessment (Radopoulou & Brilakis, 2015; Taneja et al., 2011). Imagery data can capture detailed visual information at a specified time interval. With imagery data, engineers can obtain information on the targeted civil infrastructure in both spatial and temporal domains.

The integration of automation technologies contributed to the growth of productivity in the U.S. manufacturing industry. Some researchers and practitioners expect that automation technologies such as computer vision could similarly assist in civil infrastructure operation and maintenance. Computer vision is a technique that can extract various information from imagery data such as object detection, motion tracking, and activity analysis (Cheng et al., 2013; Luo et al., 2018). In recent years, with the emergence of cost-efficient imaging sensors and the advancement of computer vision techniques, an increasing number of researchers in the domain of civil engineering began to develop and adapt computer vision techniques for civil infrastructure projects. On the

one hand, imaging sensors with different resolutions became available as well as being more cost-efficient, ranging from satellite images to digital cameras.

On the other hand, computer vision algorithms have experienced rapid development in recent years. Some studies show the potential of computer vision techniques to solve many domain challenges such as biology, aerospace, and physics. Previous research has applied computer vision to support various civil infrastructure applications with imagery data. However, large amounts of imagery data still hinder the practical use of computer vision for supporting the civil infrastructure operation and maintenance planning due to information overloading and tedious data analysis process. Therefore, the author of this dissertation proposed to use anomaly detection to assist computer vision-based civil infrastructure maintenance further to alleviate the information overloading and reduce tedious data analysis.

According to previous research, anomalies are events that deviate from the regulations, standard, or those that rarely happen and differ from the rest of the samples (Mohd Ali & Angelov, 2018). Detecting anomalies of civil infrastructure is the process of leveraging the information via inspection or monitoring processes for assessing the physical and functional conditions of civil infrastructure (Koch et al., 2015; Spencer et al., 2019). Tedious data analysis still exists after acquiring the imagery data. The author of this dissertation proposed to use anomaly detection to process the spatiotemporal details extracted from computer vision techniques. The anomaly detection algorithms can identify suspicious areas or events out of spatiotemporal information; then, the developed algorithms can highlight the targeted areas for engineers. Moreover, the developed model

integrated domain expert knowledge into the anomaly detection algorithms. The engineers can make appropriate maintenance planning by combining their expertise and suspicious areas or events highlighted by the anomaly detection algorithms.

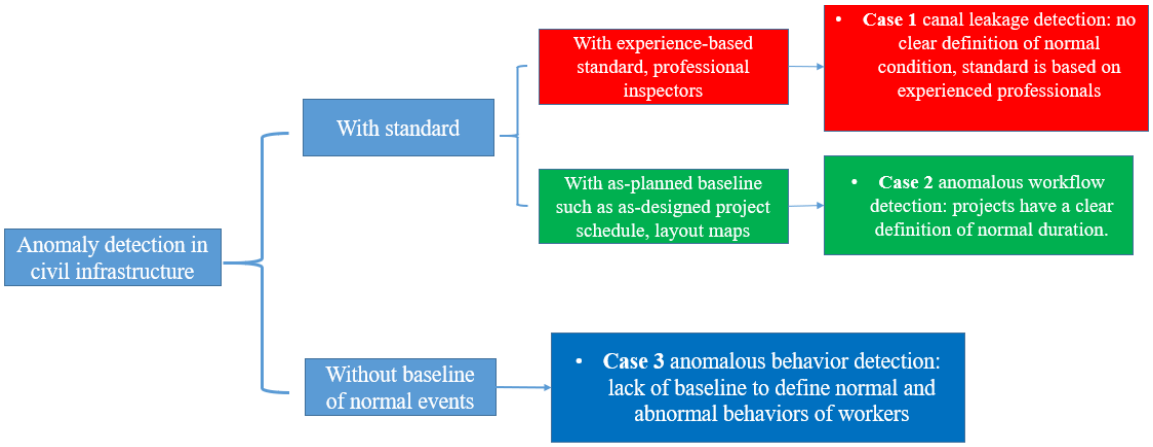


Figure 1 Categorization of Anomaly Detection in Civil Infrastructure

After reviewing and summarizing the relevant literature (Adhikari et al., 2014; Koch et al., 2015; Zaurin & Catbas, 2010), the author conducted a categorization of anomaly detection in civil infrastructure (Figure 1). Generally, there are two scenarios where engineers need anomaly detection: with and without standard.

The first scenario is that the well-established standard already exists to guide the anomaly detection. For example, in the construction project, project managers have schedule and workflow documents to make sure workers can finish the entire project on time and within budget. Moreover, for some maintenance work, there are no clearly defined baseline to distinguish the abnormal events. For example, for canal leakage inspection, the current practice is to have professional inspectors check and assess the condition of the canals. However, this approach is time-consuming and not reliable in

complex geospatial environments. Developing an efficient anomaly detection algorithm can help to mitigate the problems described above.

The second scenario is that there is no established baseline for normal events. During some civil infrastructure operations, there is no baseline to define normal behaviors. Usually, engineers identify the abnormal events when the accidents happened, which makes it challenging to have proactive control. For instance, air traffic controllers give operations to pilot and help maintain safety distance between aircraft. However, early and proactive detection of abnormal behaviors of ATC is difficult due to the lack of the standard. Chapters 3-5 give a more detailed introduction about the scenarios and how the author addressed the challenges.

The author of this dissertation proposed to use ADCV strengthened by contextual knowledge, contextual information, and contextual data to address the described challenges. To be more specific, this dissertation presented an ADCV framework. The author of this dissertation used this ADCV to examine how the proposed approaches can support anomaly detection for civil infrastructure operation in the scenarios of Figure 1 using three case studies.

The first approach is to integrate contextual knowledge into a deep learning-based water leakage detection using satellite images, which can help inspectors detect anomalous leaking sections of the canal system before maintenance activities in a cost-efficient and automated way. Through interviewing with canal inspection experts, the author of the dissertation used the canal inspection knowledge to guide the feature extraction process of deep learning algorithms. The developed approach can address the challenge that

existing algorithms cannot work in diverse geospatial environments. The second approach is a real-time multiple object tracking algorithm with the integration of contextual information to detect the anomalous workflow. This approach can mitigate the challenge that multiple object tracking algorithms cannot give a reliable performance on crowded and dynamic construction sites. The third approach is using contextual data to identify abnormal behaviors of air traffic controllers using time series of facial expressions, and head poses without a baseline of normal behaviors. This research will shed light on achieving the ADCV of civil infrastructure operation and maintenance in a cost and time-efficient way with imagery data and computer vision techniques.

1.1 Motivating Cases

This section describes the motivation of the proposed research using case studies. The chosen case studies are representative of the challenges when adopting ADCV to civil infrastructure operation and maintenance.

1.1.1 Anomalous canal section detection with the integration of contextual knowledge and deep learning

Water losses from irrigation canals amount to a significant fraction of all the water the irrigation system can deliver, resulting in problems in water conservation, soil erosion, waterlogging, and salinity (Martin & Gates, 2014). Numerous irrigation districts in western states of the US are losing a significant amount of water from their canal systems due to water leakage (Arshad et al., 2014). The United States Bureau of Reclamation reported that seepage could be reduced from 50% or more for earthen canals to 10% for

concrete-lined canals (Operation, 2017). That seepage can be further reduced to less than 5% when using a canal liner in conjunction with a concrete cover. Constructing concrete-lined canals will significantly reduce water loss through seepage.

Irrigation canals have been used for centuries to transport water to crops; the concept has remained mostly unchanged since its invention, save for the added technologies of waterwheels and hydraulics (Carter, 2015). The irrigation canal system plays a critical role in making the metropolitan area a livable community. As the oldest national multipurpose water reclamation project, the Salt River Project (SRP) operates and maintains an irrigation system that typically delivers more than 325 million gallons of water to municipal, industrial, agricultural, and urban irrigation systems annually (Salt River Project, 2015). Each fall and winter portions of SRP's 131-mile canal systems spend a month to perform maintenance work to limit water loss through seepage (Salt River Project, 2015).

In recent years, emerging techniques are showing the potential to achieve economical, fast, and precise water leakage detection, including satellite imageries and 2D/3D imaging techniques, for improved efficiency and effectiveness of canal inspection. Previous research explored various methods for the detection of water leakage, including the measurements of pressures or flow rate changes (Tucciarelli et al., 1999), acoustic signal analysis of running water (Hunaidi & Chu, 1999), and radar detection of soil moisture content (Hunaidi & Giamou, 1998). Remote sensing techniques, such as ground and satellite image data collection and processing, can automate the detection of civil infrastructures (Huang et al., 2010a). Unfortunately, such studies are still in infancy and

require engineers to manually process large amounts of remote sensory data for miles of canals and conduct field reconnaissance. A few barriers and challenges make these automation techniques impractical in field applications:

1) the geospatial environments of civil infrastructures are diverse so that no algorithms of leakage detection can handle all different environments

2) previous research used one environmental feature for leakage detection without exploring the importance of the environmental features.

The main objective of this research is to address the challenges mentioned above by integrating the canal inspection knowledge from experts. This research presents a new integration of remote sensing algorithms, domain knowledge about critical physical parameter patterns reflecting leakages and deep learning methods.

1.1.2 Anomalous workflow detection using multiple workers tracking with the integration of contextual information

Workflow surveillance is a significant aspect of determining whether workers can finish a project on time and within budget (Cheng et al., 2013; Ghanem & AbdelRazig, 2006; Girardeau-Montaut et al., 2005). Many researchers have attempted to develop a useful and timely method to detect anomalous workflow and thereby improve productivity proactively. Some researchers (Cheng et al., 2013) used the data fusion of spatial-temporal and workers' posture data to monitor workers' activity. Previous research applied many sensing and computational techniques to generate real-time data on the

location of construction entities across time, such as Radio Frequency Identification (RFID), Global Positioning Systems (GPS), and Ultra-Wideband (UWB) (Ghanem & AbdelRazig, 2006). However, all these sensors require the installation of devices on workers. Tag-based human tracking technologies are not suitable for NPP outages because NPP has restrictions on the devices that engineers can install on the job site, and trackable tasks may cause confidentiality issues (C. Zhang et al., 2017). This requirement hinders the application of these contact sensors for workspace surveillance in large-scale, congested construction sites that have many workers and objects to track.

In recent years, with the emergence of affordable video cameras and advance of computer vision techniques, an increasing number of construction companies began to set up cameras on construction sites for field surveillance. Most construction sites involve collaborative work, where there are lots of interactions and communication between workers, workers, and machines (J. Yang et al., 2010). Tracking multiple workers is thus essential to supply information to analyze how different objects interact with each other. Multiple object tracking is a computer vision technology to locate various objects, maintaining their identities, and generate trajectories of different objects given an input video (W. Luo et al., 2014). As mentioned in (Luo et al., 2018), inaccurate detection and frequent identity switch are still the major problems of multiple object tracking. The previous study did not systematically consider the identity switch in multiple object tracking, which can produce erroneous information from objects missed, mislabeled, or having discontinuities in trajectories.

The current method adopted by the construction industry has inspectors for site observations and inspection to recognize unsafe behaviors and monitor the construction workflow. Manual monitoring is time-consuming and error-prone and is not suitable for monitoring large construction sites with thousands of parallel activities (Zhu et al., 2017). Many researchers explored the potential of visual tracking to provide automated and continuous monitoring (Cheng et al., 2013). However, various challenges, including occlusions and identity switch, have been bringing uncertainties into the tracking results, and no existing studies systematically examine and quantify such uncertainties. A systematic classification and synthesis of failures of multiple object tracking methods and relevant factors that cause those failures are thus crucial for quantifying decision risks based on information derived by multiple object tracking algorithms from field videos.

Therefore, the researcher identified the primary challenge condition prognostic as crowded and rapidly changing workspaces that cause difficulties for reliable process tracking. To tackle this challenge, the researcher summarized the scenarios where multiple object tracking failed in process tracking. Next, the researcher proposed to integrate contextual information to improve the performance of process tracking. The researcher used the design map of the construction site to provide more accurate geometric information to enhance the performance of multiple object tracking.

1.1.3 Detecting anomalous behaviors of air traffic controllers in time series of facial expressions and head pose with contextual data

Air Traffic Controllers (ATCs) provide essential information and instruction to pilots, which allow pilots to maintain safe separation distances between aircraft. Facial expressions and head pose of ATCs could be indicators of changes in the ATC-pilot communication patterns and signify potential operational errors (Moon et al., 2011). For example, ATCs may miss a read-back failure of a pilot due to fatigue (Jou et al., 2013). The Federal Aviation Administration (FAA) has introduced well-designed Standard Operating Procedures (SOPs) to reduce the impacts of the anomalous behaviors of controllers (i.e., distraction, confusion) (FAA, 2005). However, human error still exists and contributes to more than 70% of all aviation accidents in the United States (U.S.). Human behavioral monitoring of ATCs is thus necessary for effectively recognizing anomalous behaviors and human error prediction during an air traffic control (Liu & Goebel, 2018).

According to the National Transportation Safety Board (NTSB), researchers identified poor ATC behaviors as one of the most significant signs that lead to human errors (Crutchfield, 2005; Nealley & Gawron, 2015). Wu also claimed that human errors caused 75% of the accidents, and fatigue related to 21% of the accidents(F. Wu et al., 2015). Sarter analyzed the NASA Aviation Safety Report System incident reports in terms of the formal characteristics of underlying errors, the cognitive stage, and the behavior at which these errors occurred (Sarter, 2009). Most incidents involved lapses (i.e., failures to perform a required action) or mistakes. Ameen identified fatigue as one of the factors that

directly affect human behavior in terms of accuracy and reaction time (Ameen, 2014). However, most of these errors were detected based on routine checks and the observed outcome of an action, respectively. Behaviors of ATCs are thus significant for ensuring the safety of the National Airspace System (NAS). An effective human behavior monitoring (e.g., fatigue detection, distraction detection, and so on) system is thus necessary to not only help detect anomalous human behaviors but also predict human errors and avoid ATC-related accidents.

Facial behaviors are the primary source of information for fatigue and distraction detection (Ameen, 2014). Facial behaviors can be natural and immediate means for human beings to communicate their emotions and intentions (Shan et al., 2009). Automatic facial expression analysis has critical applications in many areas, such as data-driven animation and human behavior monitoring (Bailenson et al., 2008). Previous studies have examined the feasibility of using computer vision techniques for human behavior monitoring through facial expression analysis (Ameen, 2014; Ba & Odobez, 2009; Ji et al., 2004; Shan et al., 2009). Soukupová has developed a real-time eye blink detection algorithm using facial landmarks to detect human operators' vigilance (e.g., driver drowsiness) (Soukupová, 2016). Reddy has developed a real-time driver drowsiness detection system by using facial landmarks of the drivers (Reddy et al., 2017).

These studies show the potential of using facial behavior analysis for human workload and fatigue monitoring. However, the main challenge is that few studies examined how to detect abnormal behaviors for diverse people in an in-time and non-intrusive manner using video and relevant data sources. There is no reference baseline of normal behaviors

for all people because everyone has their unique characteristics in their behaviors and could hardly find a unified baseline. This study developed a computer vision-based real-time facial-expression and head-pose analysis to assess the condition of ATCs. First, this study used a low-cost camera to collect real-time video to extract the head pose and facial expression of air traffic controllers. Second, this study used a long-short term memory neural network to identify the changes in the behaviors of the air traffic controllers. Last, this study also collected the heart rate data and communication data to cross-validate the anomalous behaviors identified from videos.

1.2 Problem Statement

The motivating cases showed us that anomaly detection of civil infrastructure is of importance for operation and maintenance. Although previous research proposed different approaches for anomaly detection, adopting those approaches for civil infrastructure operation and maintenance remains challenging. Compared with the scenarios used to test computer science algorithms, the construction scenarios are usually more dynamic and challenging: 1) workers have more gestures than pedestrians. 2) construction sites have extensive contextual information that could assist the computer science algorithms.

In this dissertation, the author Anomaly detection refers to two typical scenarios: 1) if there is a predefined and well-established standard, the anomaly detection algorithm can measure the as-is situation against the pre-defined criteria. 2) if there is no baseline of normal behaviors, the anomaly refers to the samples that rarely happen and differ from

the rest of the samples. How to conduct anomaly detection using computer vision approaches in civil infrastructure operation? This section summarizes the challenges of the anomaly detection of civil infrastructures as below, and the following subsections clarify the above challenges in more details:

1) Diverse geospatial environments so that no algorithms can handle the anomaly detection of civil infrastructures in all different environments.
2) The crowded and rapidly changing workspace that makes detecting anomalous task completion difficult.
3) The lack of baseline of normal behaviors so that detecting anomalous behaviors becomes difficult.

1.2.1 Challenge 1

The first challenge is that large civil infrastructures span among diverse geospatial environments so that few previous algorithms can integrate contextual knowledge to handle anomaly detection of civil infrastructures in all different environments. Establishing a baseline for normal infrastructure in different geospatial contexts is challenging. More specifically, this study chose the motivating case of canal leakage detection to demonstrate this challenge.

Significant water losses from irrigation canals can result in problems in water conservation, soil erosion, waterlogging, and salinity (Martin & Gates, 2014). Numerous irrigation districts in western states of the US are losing a significant amount of water from their canal systems due to water leakage (Arshad et al., 2014). Lining a canal with an impermeable concrete layer could help reduce such water losses. The United States

Bureau of Reclamation reported that seepage could be reduced from 50% or more for earthen canals to 10% for concrete-lined canals (Operation, 2017). A canal liner, in conjunction with a concrete cover, can further reduce the seepage to less than 5%. One problem of lined canals is that concrete cracking and deterioration could still produce water leaks. Harsh environmental conditions can erode concrete structures, and canal linings as the canals deliver water (Operation, 2017) and result in water damages and leaks. Effective monitoring and routine maintenance or repair on concrete structures and canal linings are thus paramount to keep water facilities in acceptable conditions and ensure continued safe operations of water delivery systems. Moreover, early identification of concrete deteriorations of canals could prevent expensive repairs or replacement of concrete linings later.

In recent years, remote sensing techniques are showing the potential of achieving economic, fast, and precise water leakage detection. For example, satellite imageries and 2D/3D imaging techniques have been attracting the practitioners for improved efficiency and effectiveness of the inspection of water infrastructure (Paper, 2016). Previous research explored various methods for detecting water leakage through the measurement of pressure and flow rate change (Tucciarelli et al., 1999), acoustic signal analysis of running water (Hunaidi & Chu, 1999), and radar detection of soil moisture (Hunaidi & Giamou, 1998). Remote sensing techniques, such as ground and satellite image data collection and processing, could automate the defect detection of civil infrastructures (Ozden et al., 2016).

A few barriers and challenges make these remote sensing techniques impractical in field applications: 1) Visibility challenge - most of them need direct observations of concrete and cannot assess underwater conditions; 2) Extensibility challenge - specific remote sensing and pattern classification algorithms developed for certain regions could hardly work for other areas; 3) Reliability challenge – published results indicate that most machine learning algorithms developed so far could only achieve high precision and recall for certain types of regions with relatively simple geospatial contexts (e.g., leaks in the middle of a desert), but the performance on satellite images collected from diverse geospatial environments (e.g., urban and farming areas) are either not reported or significantly weaker than simple geospatial contexts.

1.2.2 Challenge 2

In recent years, with the emergence of affordable video cameras and advance of computer vision techniques, an increasing number of construction companies began to set up cameras on construction sites for field surveillance. Most construction sites involve collaborative work, where there are lots of interactions and communication between workers, workers, and machines (J. Yang et al., 2010). Tracking multiple workers is thus essential to supply information to analyze how different objects interact with each other. Multiple object tracking is a computer vision technology to locate multiple objects, maintaining their identities, and generate trajectories of different objects given an input video (W. Luo et al., 2014). As mentioned in (Xiaochun Luo, Li, Cao, Yu, et al., 2018), severe occlusion and frequent identity switching are still the major problems of multiple

object tracking. The main challenges are the crowded and rapidly changing workspaces that cause difficulties for reliable detection of an unusual workflow. The previous study did not systematically consider the identity switch in multiple object tracking, which can produce erroneous information from objects missed, mislabeled, or having discontinuities in the tracking process.

1.2.3 Challenge 3

According to previous research, human errors account for over 70% of all aviation accidents across the U.S. Detecting the anomalous behaviors of workers such as fatigue and loss of attention can help prevent human errors and mitigate potential risks. However, the main challenge is the lack of methods for detecting anomalous behaviors for diverse people with video and contextual data sources without a baseline of normal behaviors. Air Traffic Controllers (ATCs) provide essential information and instruction to pilots, which allow pilots to maintain safe separation distances between aircraft. Facial expressions and head pose of ATCs could be indicators of changes in the ATC-pilot communication patterns and signify potential operational errors. For example, ATCs may miss a read-back failure of a pilot due to fatigue. The Federal Aviation Administration (FAA) has introduced well-designed Standard Operating Procedures (SOPs) to reduce the impacts of the anomalous behaviors of controllers (i.e., distraction, confusion) of aviation safety. However, human errors still exist and contribute to more than 70% of all aviation accidents in the United States (U.S.). Human behavioral monitoring of ATCs is thus

necessary for effectively recognizing anomalous behaviors and human error prediction during an air traffic control (Liu & Goebel, 2018).

Facial behaviors are the primary source of information for fatigue and distraction detection (Ameen, 2014). Facial behaviors can be natural and immediate means for human beings to communicate their emotions and intentions (Shan et al., 2009). Automatic facial expression analysis has critical applications in many areas, such as data-driven animation and human behavior monitoring (Bailenson et al., 2008). Previous studies have examined the feasibility of using computer vision techniques for human behavior monitoring through facial expression analysis (Ameen, 2014; Shan et al., 2009). Soukupová has developed a real-time eye blink detection algorithms using facial landmarks to detect human operators' vigilance (e.g., driver drowsiness) (Soukupová, 2016). Reddy has developed a real-time driver drowsiness detection system by using facial landmarks of the drivers (Reddy et al., 2017).

These studies show the potential of using facial behavior analysis for human workload and fatigue monitoring. However, few studies examined how time series of facial expressions and head pose of ATCs could capture the temporal patterns of ATCs' mental and physical states. There is no reference baseline of normal behaviors for all ATCs. Every ATC has his or her unique characteristics during operation and could hardly find a unified baseline for all ATCs.

1.3 Research Objectives

This proposed dissertation has the following research objectives to address the challenges mentioned above.

A. Diverse geospatial environments so that no algorithms can handle the condition prognostics of civil infrastructures in all different environments.
    - Integrate contextual knowledge to extract environmental features for canal leakage detection
    - Explore how different environmental features play a role in leakage detection in diverse environments, which can make the machine learning model explainable.

B. Crowded and rapidly changing workspaces that make anomalous workflow detection difficult.
    - Categorize and synthesize the scenarios where multiple object tracking failed in workspaces.
    - Explore the methodology of using contextual information to reduce identity switch to enable reliable detection of abnormal workflows.

C. Detection of anomalous behaviors using video and contextual data sources in an in-time and non-intrusive manner.
    - Explore the methods for supporting anomalous behavior detection from videos and contextual data without the baseline for normal behaviors.

1.4 Dissertation Organization

Chapter 0 of this dissertation provided a brief overview of the conducted research and identified the major challenges tackled in this research. This chapter used three motivating cases to illustrate the importance of using ADCV to assist infrastructure operation. Chapter 2 provided an overall description of the methodology developed in this research. Then the three chapters 3-5 are being prepared to submit for publication as journal papers. Finally, chapter 6 concludes the overall research contributions, practical implications, and highlight future research directions.

2 METHODOLOGY

This chapter presents the overall methodology developed in this research work for the ADCV. This chapter will serve to describe the importance of this research and highlight the implications. This chapter described the overall methodology to use ADCV for civil infrastructure operation. Then the author introduced the categorization of anomaly detection in civil infrastructure. Moreover, this section introduces relevant computer vision techniques. Chapters 3 - 5 are the selected case studies to demonstrate this overall methodology. These three chapters used more detailed examples and experiments to implement the developed ADCV methodology.

2.1 Vision

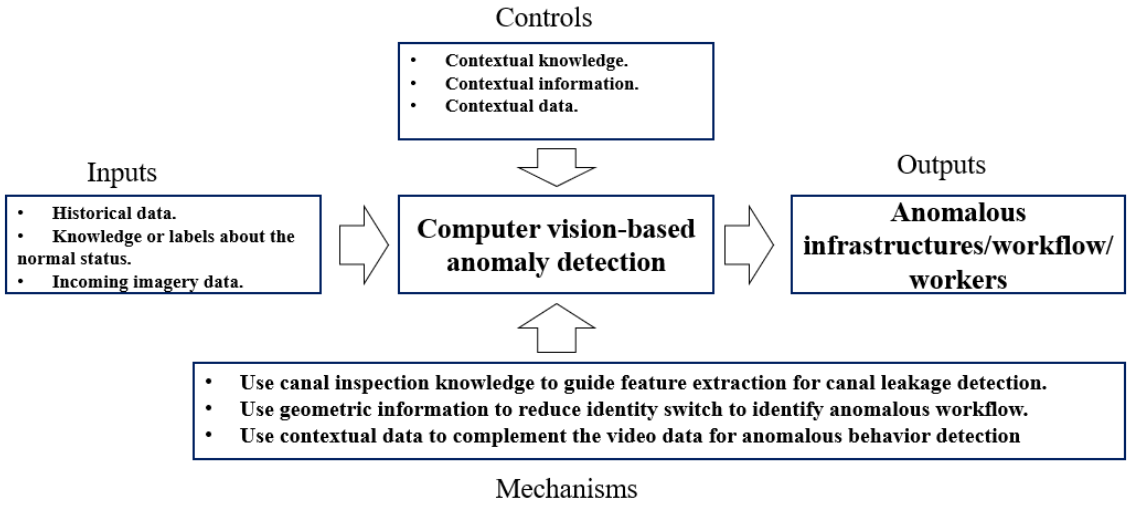


Figure 2. IDEF0 Model of Computer Vision-based Anomaly Detection (ADCV) for Civil Infrastructure Operations

The primary goals of the proposed research are using computer vision-based anomaly detection strengthened by contextual knowledge, contextual information, and contextual data to identify the anomalous infrastructure, workflow, and worker behaviors:

a) Integrate the contextual knowledge and machine learning algorithms to assist infrastructure maintenance in diverse geospatial environments.
b) Integrate contextual information to strengthen anomalous workflow detection in crowded and rapidly changing workspaces.
c) Improve the in-time and non-intrusive anomalous behavior detection of diverse people with the absence of a reference baseline of normal behaviors using video data and other contextual data.

The purpose of the proposed ADCV is to improve the efficiency and effectiveness of the anomaly detection of civil infrastructure by addressing the challenges described above. Anomaly detection of civil infrastructures is not keeping up with the recent boom of computer vision and more cost-efficient data. For example, an ideal solution to detect water leaks across a canal system should be comprehensive, efficient, precise, and economical. Previous research explored various methods for the detection of water leakages, including the measurements of pressures or flow rate changes (Tucciarelli et al., 1999), acoustic signal analysis of running water (Hunaidi & Chu, 1999), and radar detection of soil moisture content (Hunaidi & Giamou, 1998). However, these methods are either expensive or unreliable. With the emergence of satellite image and imagery sensors along with computer vision algorithms, detecting the anomalous part of the civil infrastructure has an opportunity to be automated with less cost.

2.2 Anomaly Detection in Civil Infrastructure

According to previous literature, anomaly detection is to identify the events that deviate the established standard, or those rarely happen and vary from the rest of the samples (Juba et al., 2015; Mohd Ali & Angelov, 2018). In civil infrastructure, anomaly detection can play an essential role due to the following reasons.

1) The first reason is the aging infrastructure. According to previous research(American Society of Civil Engineers, 2017), lots of civil infrastructures were built several decades ago and are well past their design life. These aging infrastructures need routine maintenance and request lots of labor and equipment resources. Moreover, according to the latest 2017 America's infrastructure report, America's infrastructure scores a D+ (ACSE, 2013). Some of the civil infrastructures need frequent monitoring of their physical and functional conditions (Kong & Frangopol, 2003). Without proper detection of the abnormal status of civil infrastructure, some accidents may happen and bring tragedies.
2) The second reason is the scale of the civil infrastructure. Many civil infrastructures have a large scale and even spanning the whole state, such as bridges, canals, pipes, and roads. Current visual inspection needs professional inspectors to conduct manual work, which is labor-intensive and time-consuming. The proposed ADCV is to identify the suspicious areas where anomalous events may happen. The inspectors can focus their efforts on the

areas targeted by the ADCV and improve the efficiency of the maintenance work.

3) The third reason is that recent technological advances in sensors, computer vision, machine learning, and other analytic techniques provide more possibility of early anomaly detection of civil infrastructure. In the past few years, different sensors are becoming more economical and accessible, which assist the anomaly detection of civil infrastructure such as satellite images for canal leakage detection and camera for worker activity monitoring.

2.3 Relevant Computer Vision Techniques for Civil Infrastructure

In recent years, with the emergence of cost-efficient imagery sensors and advanced data analytic methods, an increasing number of researchers focused on computer vision techniques. Computer vision algorithms have experienced fast development and proved the potential to tackle many domain challenges such as biology, chemistry, materials, physics, and aerospace engineering. (Basu et al., 2015; Bronstein et al., 2011; Golparvar-Fard et al., 2014). In civil engineering, some researchers began to develop and adjust computer vision techniques for civil engineering applications(Golparvar-Fard et al., 2014; Ham et al., 2016; X. Luo et al., 2018; Soltani et al., 2017; Tajeen & Zhu, 2014; Xiao & Zhu, 2018). This section listed several popular computer vision techniques that can help anomaly detection of civil infrastructure.

1) Image classification: image classification is a technique that can classify input images into the given categories. Machine learning methods can help such a

classification problem by training a classification model based on sample images and labels and applying the model to new images for the image classification.

2) Multi-object tracking: multi-object tracking (MOT) is a technique that can detect objects and generate the trajectories given video data (Bernardin & Stiefelhagen, 2008; Bewley et al., 2016; X. Li et al., 2010; W. Luo et al., 2014; Milan et al., 2016). MOT used an object detector to detect the objects in each video frame. Matching the objects across different video frames can obtain the trajectories of objects.
3) Facial feature extraction: computer vision techniques can extract human facial features, including head pose, eye blink, and facial expressions (De la Torre & Cohn, 2011; S. Li & Deng, n.d.; Werner et al., 2013; Y. Wu & Ji, n.d.). Head pose estimation refers to the ability to calculate the orientation of a person's head relative to the view of the camera. Eyeblink detection is using the facial landmark detector to calculate the eye blink rate. Facial expression recognition is to characterize the physical expression of emotions using video or image data.

3 AUGMENTING A DEEP-LEARNING ALGORITHM WITH CANAL INSPECTION KNOWLEDGE FOR RELIABLE WATER LEAK DETECTION FROM MULTISPECTRAL SATELLITE IMAGES

3.1 Introduction

Significant water losses from irrigation canals can result in problems in water conservation, soil erosion, waterlogging, and salinity (Martin & Gates, 2014). Numerous irrigation districts in western states of the US are losing significant amounts of water from canals due to water leakage (Arshad et al., 2014). Lining a canal with an impermeable concrete layer could help reduce water losses. The United States Bureau of Reclamation reported that seepage could be reduced from 50% or more for earthen canals to 10% for concrete-lined canals (Operation, 2017). A canal liner, in conjunction with a concrete cover, can further reduce that seepage to less than 5%.

One problem of lined canals is that concrete deterioration could produce water leaks. Harsh environmental conditions can erode concrete structures and canal linings (Operation, 2017) and result in water damages and leaks. Effective monitoring and routine maintenance or repair on concrete structures and canal linings are thus paramount to keep water facilities in acceptable conditions. Moreover, early identification of concrete deteriorations of canals could prevent expensive repairs or replacement of concrete linings later.

Concrete deterioration assessment of canals and water leak analysis also has significant socio-economic impacts. Irrigation canals have been used for centuries to transport water to crops(Carter, 2015). The irrigation canal system plays a critical role in making the

metropolitan area a livable community. For example, as the national oldest multipurpose water reclamation project, the Salt River Project (SRP) operates and maintains an irrigation system that typically delivers more than 325 million gallons of water to municipal, industrial, agricultural, and urban irrigation systems annually (Salt River Project 2015). Each fall and winter, SRP personnel spend around two months to perform maintenance to reduce seepage on selected sections of the 131-mile canal network in Phoenix (Salt River Project, 2015). The selection of the canal sections requires inspectors to assess the deterioration trends of the concrete lining of canals for effective maintenance planning.

The current approach of canal inspection often involves manual inspection with limited technologies for detailed and objective condition assessments. In many cases, even spending hours in the field, inspectors could still miss defects and could hardly measure sediments and defects hidden under the water or buried in the soil. According to the manual of *Canal Operation and Maintenance: Concrete Lining and Structures* by U.S. Department of the Interior, Bureau of Reclamation (Interior, 2017), establishing a formal inspection program is vital to maintaining the condition of concrete structures and canal linings. Inspectors must identify and assess the signs of the surface damages meticulously. Specifically, the challenge of canal condition assessment lies in the tedious canal condition assessments that mostly produce field notes that are subjective and could hardly support reliable deterioration trend analysis of concrete lining.

In recent years, remote sensing techniques are showing the potential of achieving economic, fast, and precise water leakage detection. For example, satellite images and

2D/3D imaging techniques have been attracting the practitioners for improved efficiency and effectiveness of the inspection of water infrastructure (Huang et al., 2010b). Previous research explored various methods for detecting water leakage through measuring pressure and flow rate changes (Tucciarelli et al., 1999), analyzing acoustic signals of running water (Hunaidi & Chu, 1999), and using radar to detect the soil moisture (Hunaidi & Giamou, 1998). Remote sensing techniques, such as ground and satellite image analyses, could automate the assessment of the conditions of civil infrastructures (J. Chen et al., 2017; Huang et al., 2010b). A few challenges make these remote sensing techniques impractical: 1) **Visibility challenge -** most of them need direct observations of concrete and cannot assess underwater conditions; 2) **Extensibility challenge -** certain remote sensing and pattern classification algorithms developed for certain regions could hardly work for other regions; 3) **Reliability challenge** – most machine learning algorithms published achieved high precisions and recalls in relatively simple environments (e.g., leaks in the middle of a desert), but the performance on satellite images collected from diverse environments (e.g., urban areas) are either not published or significantly weaker than simple contexts.

This research aims at addressing the challenges mentioned above through a deep-learning approach augmented by domain knowledge about particular physical parameter spatiotemporal patterns that possibly reflect leaks. For example, thermal dynamics knowledge about how water penetration influences the temperature distributions around leaking canals could guide engineers to find leaks. Machine learning approaches augmented by such physics-based knowledge are "*Physics-based learning methods*."

Some researchers augmented artificial neural networks, a type of machine learning methods, with physics-based domain knowledge, and call such techniques as “Physics-Guided Neural Networks” or PGNN (Karpatne et al., 2017). Unlike the conventional “black-box” neural networks, this PGNN approach could help researchers examine and explain how different environmental conditions influence the performance of the developed machine learning models in leak detection. Specifically, the authors expect that the PGNN approach could achieve improved precision and recall rates in water leak detection on satellite images collected from diverse geospatial environments. With that in mind, the researchers established a computational framework to validate the performance of a PGNN approach in water leak detection based on satellite images.

Compared with machine learning algorithms trained by labeled raw satellite image samples, the new deep learning algorithms augmented by canal inspection knowledge can use satellite images augmented by pixel-level land surface temperature (LST), fractional vegetation coverage (FVC) and Temperature Vegetation Dryness Index (TVDI) as training samples of leaking and non-leaking canal sections. More specifically, LST, FVC, and TVDI for each pixel are physical parameters derived from Landsat 8 satellite images by remote sensing algorithms. Literature review results and domain knowledge of canal inspectors both indicate that specific patterns of these physical parameters can serve as reliable indicators of water leaks. The proposed method uses the augmented satellite image samples with pixel-level LST, FVC, and TVDI values to train a deep learning network for classifying leaking and no-leaking sections of canals. The researchers collected leakage data of canal systems in Arizona, the US from the year 2016 – 2018

annually to test the developed methodology in both urban and rural contexts. The collected data included the geolocations of the canal leakage and Landsat 8 multispectral satellite images. Furthermore, the authors explored different combinations of physical parameters to identify feature combinations that make the new deep learning approach achieve better accuracies, precisions, and recalls in leak detection.

The organization of the remaining sections of this paper is as follows. Section 3.2 reviews previous studies related to canal leakage detection based on satellite imagery data. Section 3.3 describes the technical details of the Convolutional Neural Network (CNN)-based leakage detection algorithm augmented by the inspection knowledge. Section 3.4 presents the experiment design for validation. Section 3.5 presents testing results to show the performance of the developed algorithm regarding accuracy, precision, and recall using images collected from diverse environments. Moreover, the authors compared the developed algorithm with the conventional deep learning algorithm that uses raw images for training. Section 3.6 concludes the study and discusses the limitations of the developed algorithm and future research.

3.2 Previous Research

This section synthesizes studies relevant to the use of remote sensing and machine learning techniques for water leak detection in multiple domains. Section 3.2.1 summarizes relevant studies on canal leakage detection methods. Section 3.2.2 outlines the use of Physics-guided Neural Networks (PGNN) in various applications to show the potential of the physics-guided machine-learning approach for improving the image-based leak detections.

3.2.1 Canal leakage detection

An ideal solution to detecting water leaks across a canal system should be comprehensive, efficient, precise, and economical. Previous studies explored the use of sonar imagery and remotely operated vehicles or autonomous underwater vehicles in supporting underwater inspections (Inzartsev & Pavi, 2009). However, these approaches require professional surveyors to collect data on-site. The efficiency and costs for the data collection become the bottlenecks that limit the wide adoption of these approaches in large canal networks that contain hundreds of miles of canals. Previous researchers have identified that remote sensing has the potential for efficient and effective water leakage detection (Hadjimitsis et al. 2011; Huang et al. 2010). The uses of remote sensing techniques for water leakage detection are time- and cost-effective compared with traditional, intrusive methods such as acoustic sensors (Martin & Gates, 2014). Previous studies have proven that vegetation on the levee, the temperature in the surroundings, soil humidity are common indicators of canal leakages (Saha, 2015).

Huang et al. developed an airborne system mounted with red, near-infrared, and thermal sensors to collect multispectral images (Huang et al. 2010). Combined with field reconnaissance, the researchers manually rate sections of canals in terms of wetness, grass growth, cracks on the levee. This method could identify leaking parts, but it is subjective. When inspectors are not familiar with particular environments and canal sections, the manual assessments are unreliable.

Zanganeh et al. used Landsat 8 satellite image processing techniques to locate leakages (Zanganeh 2016). They showed that remote sensing techniques using free medium

resolution satellite images could achieve early detection of leakages. The developed algorithm used a normalized difference vegetation indices (NDVI) to analyze the distribution of vegetation along the canals. The algorithm then used the K-means classification technique on the NDVI map to identify leakages. However, this method only considered the NDVI, while NDVI could hardly provide sufficient information for leakage detection in environments that mix buildings, plants, and other land covers. Besides, the K-means method needs the users to input a K value, but that K value estimated by the users could be inaccurate and misleading the clustering results. Although sensitivity analysis can help to find the optimal K values, the optimal K values could vary with different data sets.

The methods mentioned above are either subjective or requiring prior knowledge crafted by domain experts. Personal factors pose challenges to objective and consistent condition assessment of canal network spanning over in various geospatial environments – even the most experienced experts could not be familiar with all sections of canals in all possible contexts. Moreover, these methods focused on one or two environmental features (vegetation, temperature, soil humidity), while pointing out that none of those features alone could support robust, consistent, and reliable leak detection in various environments (Huang et al., 2010a; Zanganeh et al., 2016).

3.2.2 PGNN

Previous researchers have adopted computer vision, and machine learning approaches for civil infrastructure inspection tasks, such as pavement defect detection (Koch et al., 2015; Radopoulou et al., 2016). However, those approaches cannot work well in canal crack

detection because the cracks of canal systems are usually under the water and invisible. Moreover, the conventional “black-box” neural network has one limitation - the trained model is solely dependent on the training data. In many cases, the predictions of the model could be inconsistent with the known laws of physics. Recent research proposed two categories of PGNN. The first category of PGNN uses physics theories to calculate and feeds features into neural networks. Karpatne et al. proposed PGNN and applied the methodology in lake temperature modeling (Karpatne et al., 2017). That study integrated the physical relationships between the temperature, density, and depth of water into the PGNN

The second category of PGNN adopts physics-based loss functions by adding a physical-inconsistency term. For example, Yu et al. adopted a PGNN approach for aircraft dynamics simulation (Y. Yu et al., 2019). The researchers used the underlying physics of the aircraft dynamical system to construct a deep residual recurrent neural network. The results showed that PGNN has better generalization potential and could produce physically meaningful results to improve the interpretability of the neural networks (Karpatne et al., 2017; Y. Yu et al., 2019).

3.3 Research Methodology

The methodology developed by the authors consists of three parts (

Figure 3). Firstly, the research team developed a new method to utilize satellite images to evaluate the condition of canals by including more environmental features: land surface temperature (LST), Fractional vegetation cover (FVC), and soil moisture (TVDI). The

purpose is to address the fact that existing remote sensing studies for water leakage detection had limited testing and validation in complex urban environments while mainly relying on Normalized Difference Vegetation Index (NDVI). Secondly, the researchers applied a convolutional neural network to classify the satellite images as a leaking section and no-leaking sections based on historical canal maintenance records. The third step is the post-processing of the classification results of CNN. The classification algorithm predicts the leakages by separating the large satellite images into small windows. This step needs to geo-reference the small windows to locate these windows on the satellite images and help determine the leaking sections within those windows. More detailed illustrations of each step are as follows.

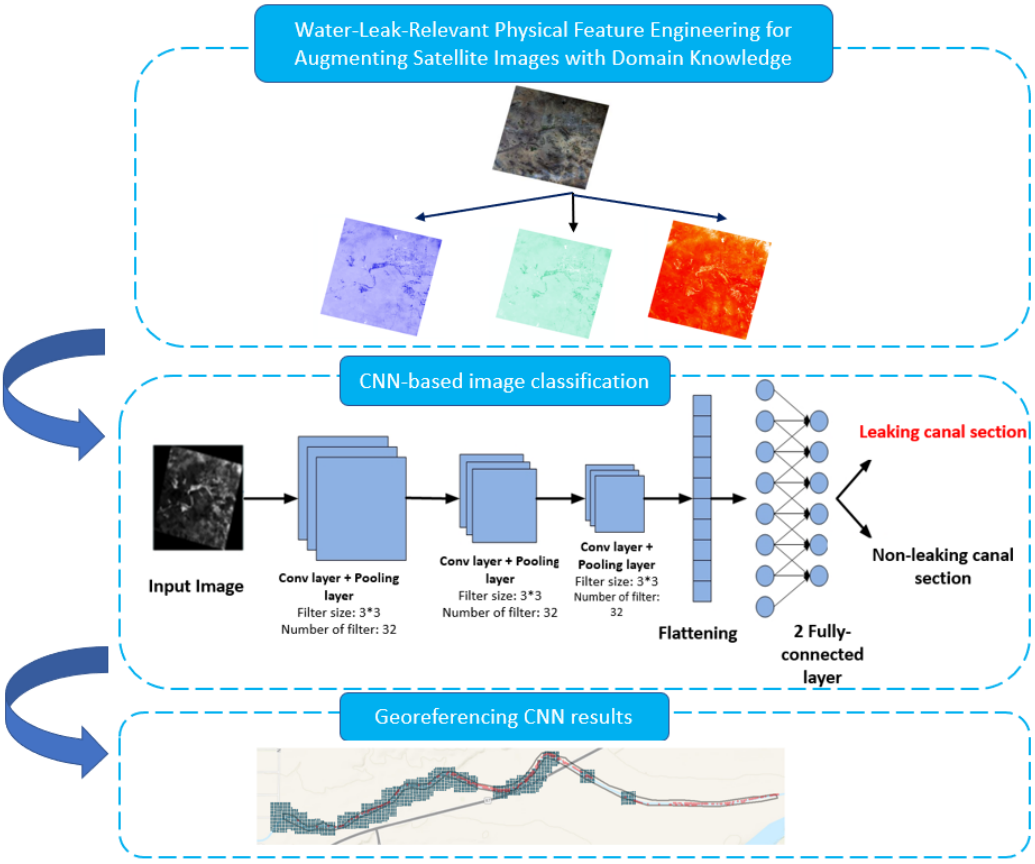


Figure 3. An Architecture for Classification of Canal Conditions Using Satellite Images

3.3.1 Water-leak-relevant physical feature engineering for augmenting satellite images with domain knowledge

The features in the input data of deep learning are important for the prediction. The quality and quantity of the features will have significant impacts on the model's classification performance. Based on the domain knowledge of irrigation canal management, LST, FVC, and TVDI tended to be reliable indicators of canal leakages (Arshad et al., 2014; Operation, 2017). For instance, if vegetation appears in a dry area where the plant is uncommon, it indicates that seepage may exist.

Similarly, water loss arisen from canal leakage influences the soil humidity and land surface temperature of the surrounding areas (Huang et al., 2010a). Some researchers used radar to detect soil humidity and used thermal cameras to detect land surface temperature for canal leakage detection (Nellis, 1982). The authors used the Landsat 8 satellite images to derive LST, FVC, and TVDI as the input features of the developed PGNN algorithm.

LST is a crucial indicator of water leakage (Huang et al., 2008). The researchers used an approach called the Radiative Transfer Equation (RTE) to calculate the LST from high-resolution remote sensing images (X. Yu et al., 2014). RTE represents the propagation of electromagnetic radiation through the earth's space. The processes affecting that propagation include absorption, emission, and scattering. Yu and his colleagues proposed a simple version radiative transfer equation expressed as follows (X. Yu et al., 2014),

$$R_{i,T_i} = \tau_{i,\theta} \times \left(\varepsilon_i \times R_{i,G} + (1 - \varepsilon_i) \times PR_{i,\downarrow}\right) + PR_{i,\uparrow} \qquad \textit{Equation 1}$$

where R_{i,T_i} refers to the sensor radiance of channel i based on the brightness temperature T_i. $R_{i,G}$ refers to the ground radiance. $PR_{i,\downarrow}$ and $PR_{i,\uparrow}$ refers to the downwelling and upwelling path radiance, respectively. $\tau_{i,\theta}$ refers to the atmospheric transmittance of channel i within the zenith angle θ. ε_i refers to the surface emissivity of channel i.

Moreover, $R_{i,G}$ is expressed by the law of Plank,

$$R_{i,G} = 2 \times ac^2/(\mu_i^5 \times (e^{ac/\mu_i bT_s} - 1)) \qquad \textit{Equation 2}$$

where a, b and c refer to the light speed, Planck constant, and the Boltzmann constant, respectively. μ_i is the wavelength of channel i.

$$T_s = \frac{C_1}{\mu_i \, ln(\frac{C_2}{\mu_i^5 (R_{i,T_i} - PR_{i,\uparrow} - \tau_i(1-\varepsilon_i)PR_{i,\downarrow})/\tau_i\varepsilon_i} + 1)} \qquad \textit{Equation 3}$$

Where $C_1 = 14387.7\ \mu m \cdot K$, $C_2 = 1.19104 \times 10^8 W \cdot \mu m^4 \cdot m^{-2} \cdot sr^{-1}$.

T_s is the land surface temperature. The authors could estimate $PR_{i,\uparrow}$, $PR_{i,\downarrow}$ and τ_i from the radiative transfer model with the thermal radiance measured at the sensor level and the atmospheric parameters obtained with the radio sounding. Following Equation 3, the authors can derive LST.

Then, the researchers used the Temperature Vegetation Dryness Index (TVDI) for soil humidity estimation. TVDI is a parameter to normalize the surface soil moisture. The TVDI method combines visible, infrared, and thermal bands. Based on the previous research, $TVDI = (T_{st} - T_{min-st})/(T_{max-st} - T_{min-st})$ *Equation 4* calculates TVDI (Younis & Iqbal, 2015).

$$TVDI = (T_{st} - T_{min-st})/(T_{max-st} - T_{min-st})$$ *Equation 4*

where T_{st} refers to the surface temperature from Landsat 8. T_{min-st} and T_{max-st} denotes the minimum surface temperature and maximum surface temperature, respectively. the following equations calculate T_{max-st} and T_{min-st} :

$$\begin{cases} T_{min-st} = a_1 + b_1 \times NDVI \\ T_{max-st} = a_2 + b_2 \times NDVI \end{cases}$$ *Equation 5*

where a_1 and b_1 are the coefficients used for controlling T_{min-st}, and a_2 and b_2 are the coefficients used for controlling T_{max-st} . $NDVI = (w_{if} - w_r)/(w_{if} + w_r)$ *Equation 6* calculated $NDVI$.

$$NDVI = (w_{if} - w_r)/(w_{if} + w_r)$$ *Equation 6*

where w_{if} and w_r refer to the near-infrared and red wavebands, respectively.

FVC is the ratio between the vertically projected area of vegetation and the total surface extent (Song et al., 2017). FVC is widely used to describe vegetation quality. The authors implemented the linear mixture model, which has been widely applied for the estimation of FVC to generate vegetation coverage (Gutman & Ignatov, 1998).

$$FVC = \frac{NDVI - NDVI_s}{NDVI_v - NDVI_s}$$ *Equation 7*

Where $NDVI_s$ *is the minimum of* NDVI in the studied area and $NDVI_v$ is the maximum of NDVI in the studied area (Gutman & Ignatov, 1998).

3.3.2 CNN-based image classification

The water leakage detection based on satellite image analysis is an image-based classification problem of regions captured in satellite images. In other words, the algorithm developed by the researchers take satellite images as inputs and classify the areas of the photos as leaking or non-leaking sections. Machine learning methods can solve such a classification problem by training a classification model based on sample images and applying the model to new images for the image region classification.

One category of the machine learning algorithm is the artificial neural networks (ANNs) (Mitra et al., 2006), which is the basis of the more recently developed deep-learning network algorithms (L. Zhang et al., 2016). Rather than sending surveyors to the canals and conduct inspections, the machine learning algorithm can take large amounts of image samples and learn from the examples to estimate a set of parameters for a neural network model. The resulted neural network can classify new images based on the training samples. This parameter estimation process of the ANN model is the "training" process. The trained ANN model uses the similarity between training samples and images for classification. Deep learning is a new generation of machine learning algorithms after ANN. Deep learning simulates the human brain, which consists of many layers of neurons allowing it to make complex decisions. In this research, the researchers used historical maintenance data to train a deep learning model. The deep learning algorithm can achieve more reliable classifications, given large amounts of historical image samples.

The research team extracted feature images of the LST, FVC, and TVDI from satellite images. With these three feature images, the research team reconstructed a three-channel

image as the input of the convolutional neural network. Using a sliding window technique, the researchers segmented the satellite image into 8*8 windows with a 1-pixel step of moving the window on satellite images. Figure 4 shows the architecture of the model. This model includes three convolutional layers and two fully connected layers. In this model, the research team conducted a binary classification to categorize image windows. The output is the condition of each small window being leaking or no-leaking (e.g., canal section has no leakages, or the canal section has leakages).

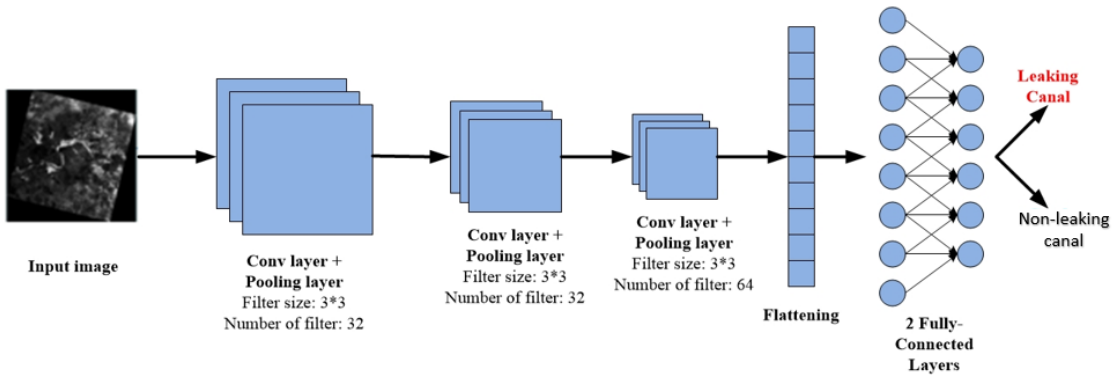


Figure 4. Convolutional Neural Network Architecture.

The architecture of the proposed CNN consists of the conv-pooling layer, the fully-connected layer, and the classification layer. The input image has dimensionality 8*8*3, and the three channels are LST, FVC, and TVDI. The researchers resized the image to 128*128*3 to feed into the CNN. The conv-pooling layer includes three sub-layers; the first and second conv-pooling layers use 32 filters within the 3*3 window. The last conv-pooling layer uses 64 filters within the 3*3 window. Then a two-layer fully connected layer processes the output of flattening the result generated by the last conv-pooling layer.

Finally, the classification layer uses a "softmax" function to perform a binary classification to determine whether each image belongs to no-leaking or leaking condition. "Softmax" is a function that can output a vector, which represents the probability distribution over the predicted output classes (Krizhevsky et al., 2012). The predicted class is the class with the highest probability output from the softmax function.

After collecting the satellite image and locations of repairs, the next step was to segment the satellite image into multiple 8*8 windows and annotate the windows. The annotation procedure is different from traditional computer vision annotation that visually labeled every image based on subjective understanding. In this project, the locations of repairs are labeled on the Google Earth satellite image first, followed by overlaying the repair locations on the Landsat 8 images. This overplayed product helps the research team identify the pixels which crossed the physically repaired sections of canals automatically rather than labeling each image. If the 6*6 square (red square in Figure 5) in the 8*8 window contains a repair location, then the research team labeled the window as a window containing leaking sections (called "leaking window" hereafter), and vice versa. Such labeling is more objective and reflecting the actual repairing needs for leaking parts.

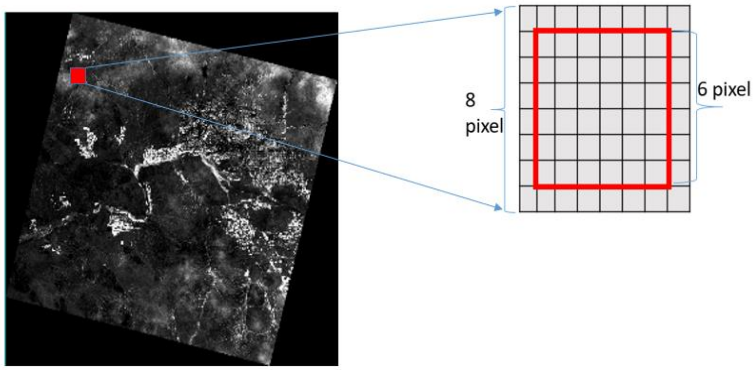


Figure 5 Annotation of 8*8 Windows Extracted from Satellite Imagery Data

3.3.3 Georeferencing the results generated by the CNN algorithm

The traditional CNN algorithm gives the prediction result of each image sent to the algorithm. CNN predicts the conditions of every window being “no-leaking” or “leaking.” However, different from the other imagery data, the windows generated from satellite images are geo-referenced (L. Zhang et al., 2016). The research team projected each window back to the map to locate which sections of canals are having leaking parts. For this purpose, the research team used ArcGIS Desktop to geo-reference the windows.

3.3.4 Performance metrics for evaluating the developed PGNN approach

This section presents the performance evaluation approach used to quantify the performance of the developed PGNN classification algorithm. Computer science researchers defined the following metrics to measure the performance of machine learning algorithms. The research team thus uses these metrics for quantitatively assessing the classification performance of the developed algorithm.

- True positive (TP) means that the algorithm indicates the presence of a condition when, in reality, it is present.
- True negative (TN) means that the algorithm indicates no presence of a condition when, in reality, it is not present.
- A false positive (FP) is an error in which the algorithm indicates the presence of a condition when, in reality, it is not present.
- A false negative (FN) is an error in which the algorithm indicates no presence of a condition when, in reality, it is present.

The evaluation generates four values, including true positive, false positive, false negative, and true negative. For example, "true positive" means that the CNN algorithm predicted the canal section had cracks, and the section has cracks in reality. In this case, if the SRP team repairs the section predicted by the CNN algorithm, they would be able to allocate the resources correctly. *Equation 8* defines the precision of the algorithms. Precision means that, for all the leakage sections detected by the algorithm, the algorithm correctly detected portions of the leakage sections. The recall represents that, among all the leakage sections, in reality, the percentages of the leakage sections the algorithm correctly detected.

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \; Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 8 Precision, recall, and accuracy of the algorithm

3.4 Experiments

The researchers used the past maintenance records of the canal sections to predict the water leakage of other canal sections that engineers need to maintain in the future. This

section describes the studied areas the researchers chose for the experiments and the data preparation process for the tests.

3.4.1 Studied areas

Arizona Canal, South Canal, and Western Canal are three main canals belonging to the SRP canal system. Arizona Canal and South Canal crossed rural areas, while the Western Canal crossed urban areas. The authors chose these areas to compare the performance of the algorithm in different environments. Figure 6 shows the study areas in this experiment. SRP dried-up Western canal in December 2016, South canal around December 2017, and Arizona Canal in January 2018.

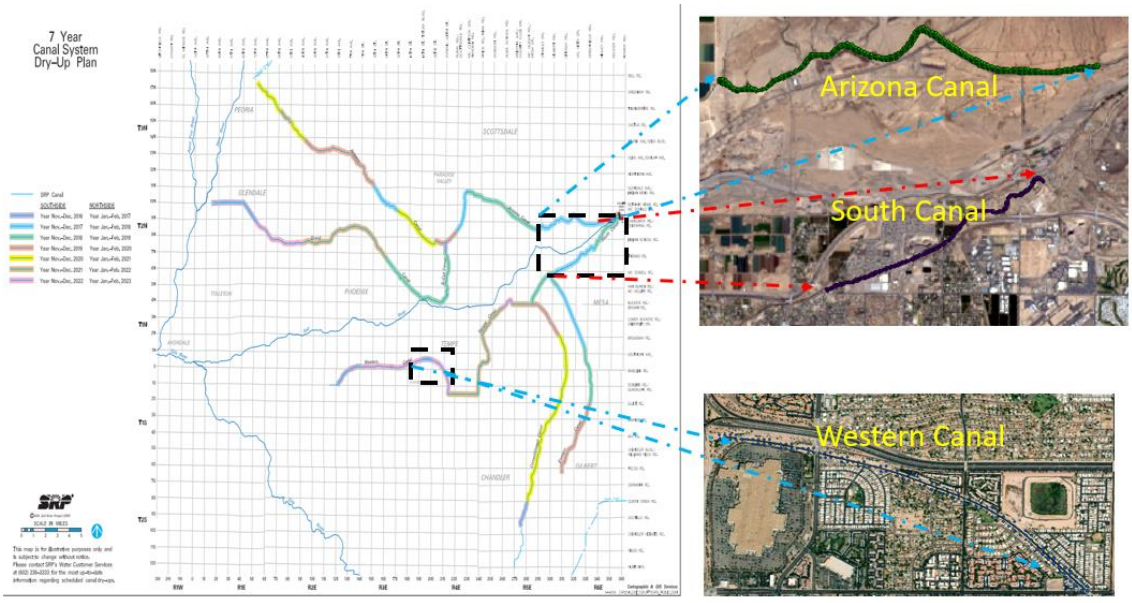


Figure 6. Study Areas. Arizona Canal and South Canal are in Rural Areas. Western Canal is in Urban Areas.

3.4.2 Data preparation

The researchers chose Landsat 8 satellite images, which are free to the public since 2013. The Landsat 8 satellite can take images of the entire earth every 16 days (Zanganeh 2016). Landsat 8 has 11 bands, multispectral bands 1-7 and 9 (30-meter pixel size), panchromatic band 8 (15-meter pixel size), and thermal infrared bands 10-11. Based on the algorithms developed in previous studies, the research team extracted environmental features, including LST, FVC, and TVDI (Gao et al. 2011; Younis and Iqbal 2015; Yu et al. 2014).

After the dry up and repair of the South Canal, the research team used the repair information provided by SRP to locate the concrete cracks for repairing (Figure 7) and labeled the locations of leakage based on the maintenance records using ArcGIS (Figure 8). ArcGIS can give the leakage locations geocoordinates for alignment with the coordinate system used by satellite images. The sliding-window techniques then segment the satellite images into the 8*8 window (64 pixels each window) with a step at 1 pixel. The authors filtered out the windows that do not cross the canal and then labeled the windows that cross the canal as leaking and non-leaking. Finally, the authors collected a dataset of 6,360 windows with 4,409 windows as leaking and 1,951 windows as non-leaking (Table 1). The authors made different transformations, including flip and rotation, to the original images to increase the training dataset. Such augmentation of the training data set has the name “Data augmentation” in the literature (Shorten & Khoshgoftaar, 2019). The augmented dataset has 10,000 windows. Finally, the researchers used the

expanded dataset to train and test the proposed method with 8,000 windows for training and 2,000 windows for testing.

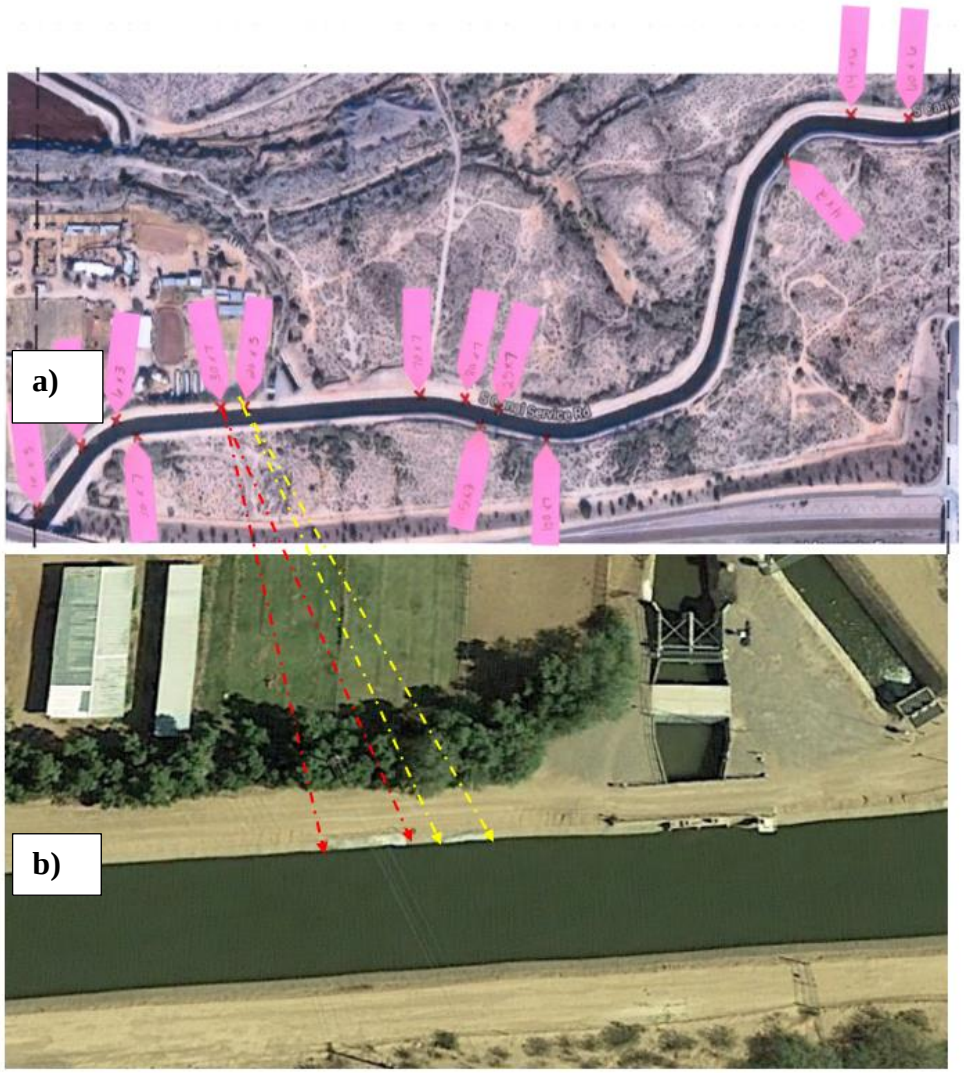


Figure 7. (a) An Example of the Maintenance Records that Highlight the Dimensions and Locations of Leaking Parts of the Canal. (b) the Actual Location of Leakage Marked by Surveyors on Canal Riverbed, Image from Google Earth.

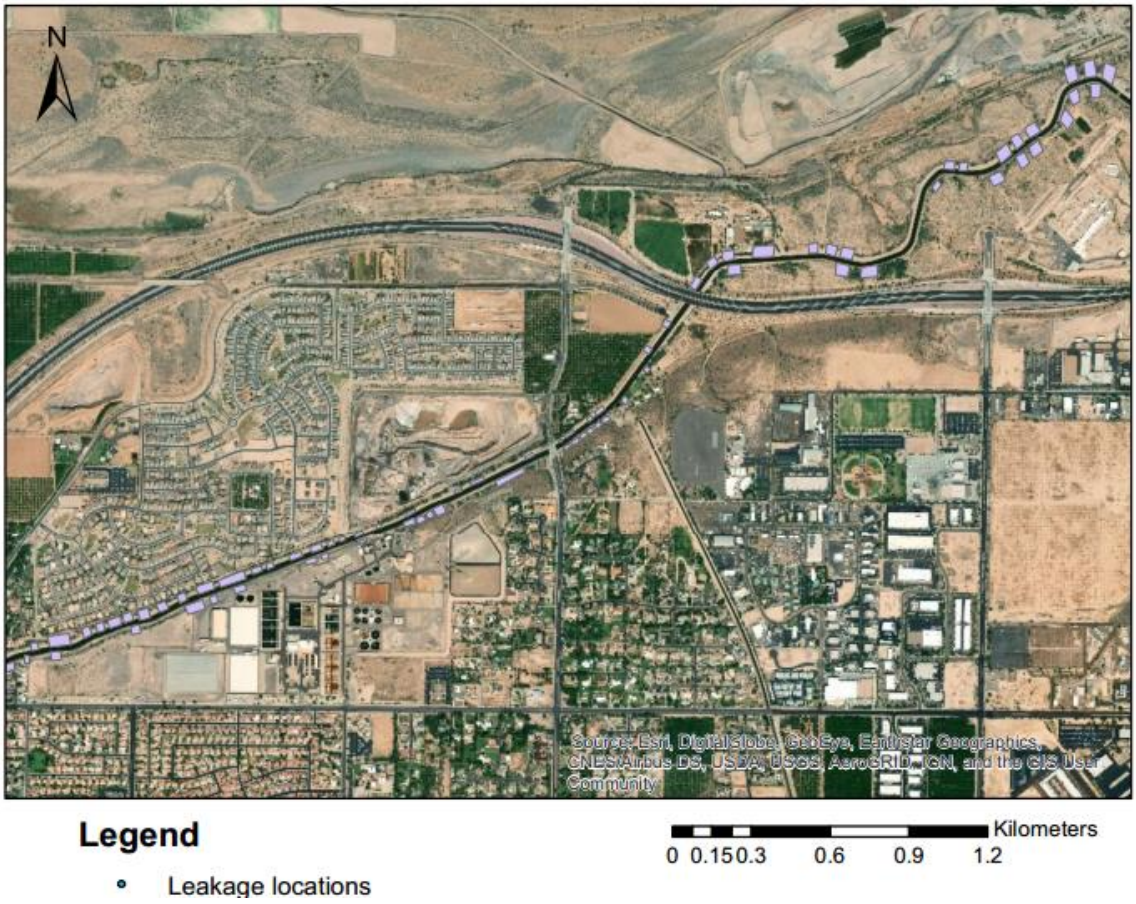


Figure 8. Mapping the Leakage Locations Using the ArcGIS Platform: The Purple Boxes of Different Sizes Indicate the Dimensions and Areas of the Leakages.

Table 1 Data Description

Canal	Area type	Satellite image date	Number of leaking windows	Number of non-leaking windows
Western Canal	Urban	11/01/2016	179	698
South Canal	Rural	10/19/2017	1684	271
Arizona Canal	Rural	01/23/2018	2546	982

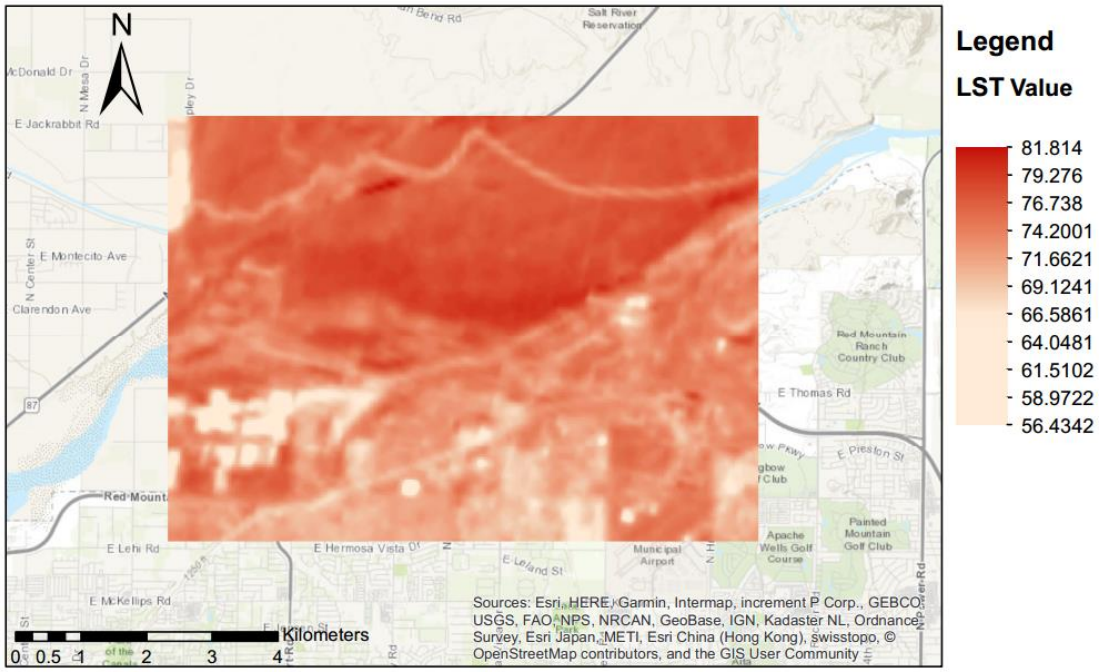
As for the validation part, the research team marked the repair location of the Arizona Canal on the high-resolution satellite image on Google earth shown in Figure 8.

3.5 Results

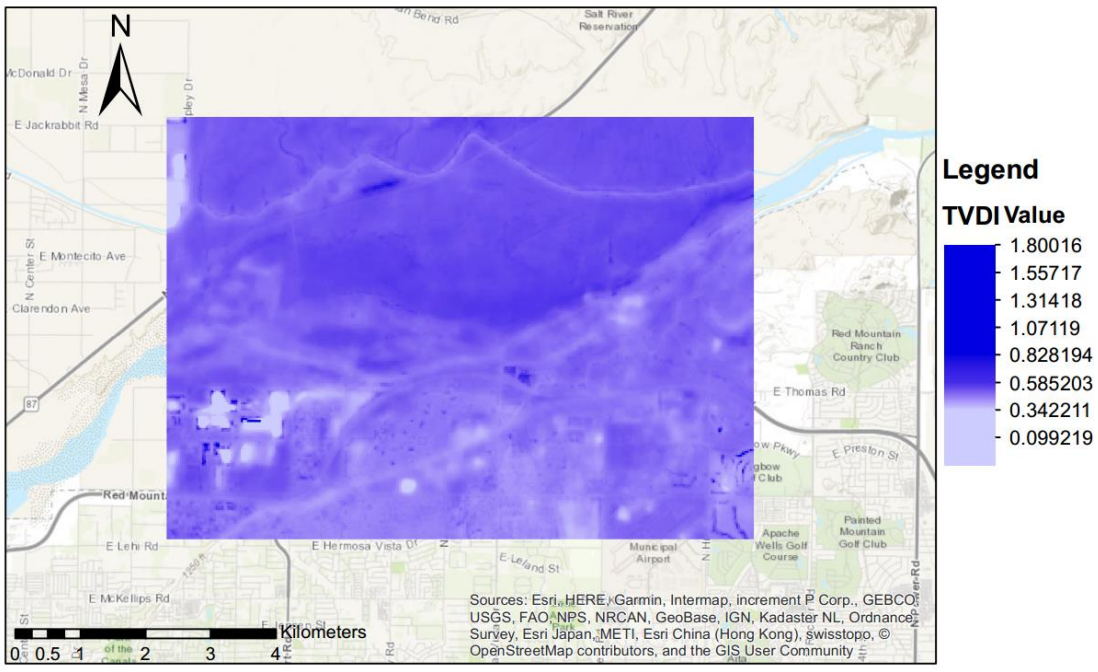
This section presented the results of the experiments and some discussions about the findings. The researchers extracted the LST, TVDI, and FVC from satellite images and detected water leakages using the proposed methodology. To further explain the benefits that this methodology could bring to the canal management, the researchers compared the current canal maintenance practice against the new methodology. Finally, the researchers tested the performance of the algorithm on different landcover: urban and rural to evaluate the performance of the algorithm.

3.5.1 Environmental feature extraction results from satellite images

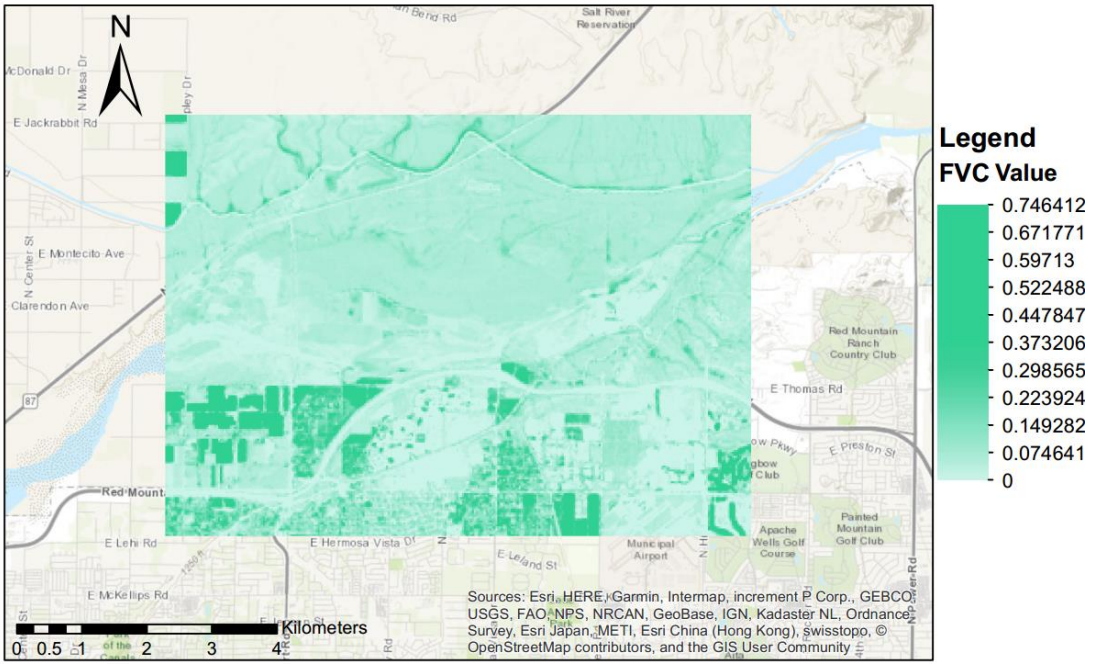
Following the feature extraction methods described in Section 3.3.1, the researchers extracted the LST, FVC, and TVDI from satellite images. The Landsat 8 image has a resolution of 30 meters, which means every pixel in the image represents 30 meters. Figure 9 shows the results of deriving LST, FVC, and TVDI using Landsat 8. The higher value of the pixel in the plant cover image represents more vegetation in that area. Similarly, the more significant value of the pixel in the land surface temperature result represents a higher temperature in that area. However, the smaller value of the pixel in soil humidity image represents more soil.


N
Legend
LST Value
81.814
79.276
76.738
74.2001
71.6621
69.1241
66.5861
64.0481
61.5102
58.9722
56.4342
Salt River Reservation
McDonald Dr
N Mesa Dr
E Jackrabbit Rd
E Montecito Ave
N Center St
Clarendon Ave
Red Mountain Ranch Country Club
E Thomas Rd
E Preston St
E Lehi Rd
E Hermosa Vista Dr
Municipal Airport
Apache Wells Golf Course
Painted Mountain Golf Club
E Leland St
E McKellips Rd
Kilometers
0 0.5 1 2 3 4
Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community

(a)


N
Legend
TVDI Value
1.80016
1.55717
1.31418
1.07119
0.828194
0.585203
0.342211
0.099219
Salt River Reservation
McDonald Dr
N Mesa Dr
E Jackrabbit Rd
E Montecito Ave
N Center St
Clarendon Ave
Red Mountain Ranch Country Club
E Thomas Rd
E Preston St
E Lehi Rd
E Hermosa Vista Dr
Municipal Airport
Apache Wells Golf Course
Painted Mountain Golf Club
E Leland St
E McKellips Rd
Kilometers
0 0.5 1 2 3 4
Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, © OpenStreetMap contributors, and the GIS User Community

(b)



(c)

Figure 9 a: Land Surface Temperature (LST), b: Soil Humidity (TVDI) and c: Vegetation Coverage (FVC) Results in the Studied Area

3.5.2 Leakage detection results using CNN

The researchers evaluate the performance of the proposed algorithm in terms of accuracy, precision, and recall (*Equation 8*). The researchers used the collected dataset (Section 3.4.2) that includes canals in different landcover from the year 2016 to the year 2019. The researchers used all three environmental features LST, FVC, and SH, to train the CNN model. The precision of the methodology is 86%, the recall is 86%, and the accuracy is 85%.

Moreover, the researchers tested the performance of the trained model in different environments of different land covers. Figure 10 showed the recall, accuracy, and precision

of the same trained model tested in urban and rural areas. The researchers found that the trained model has a more reliable performance in rural areas. This finding coincides with the fact that urban areas have more complex geospatial environments that make urban areas more challenging than rural areas to detect canal leakages.

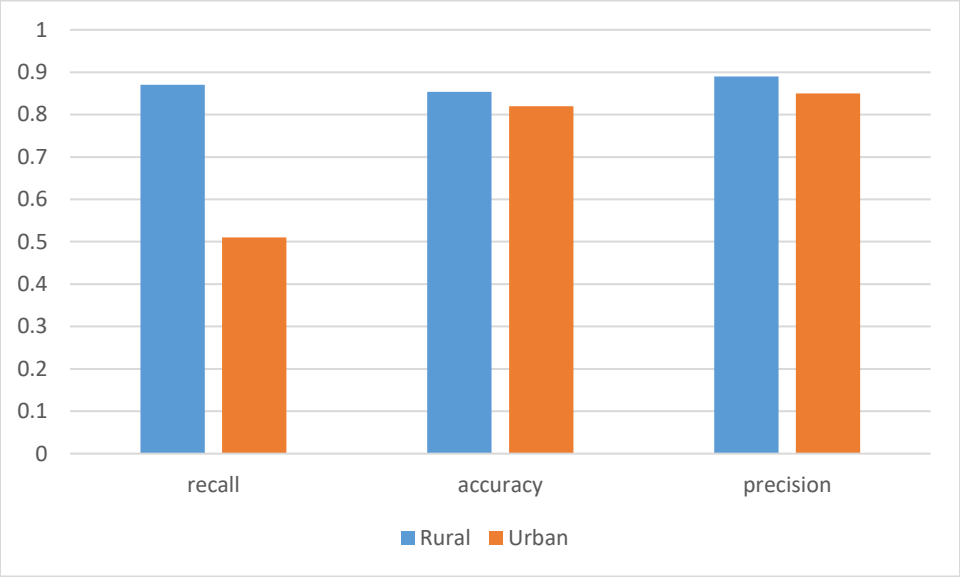


Figure 10 Comparison of the Performance in Different Environments

3.5.3 Test on environmental feature combination

One major limitation of the neural network is that the neural network is similar to the "black box." The trained model can establish a relationship between input and output, using millions of parameters. However, the neural network cannot tell what happened inside the "black box" and which parts of the input are more informative to the output. Knowing which features are more influential to canal leakage detection is essential to engineers. The research team proposed to use different combinations of environmental features as training data and test the performance of the proposed algorithm in both rural and urban areas.

Because there are three environmental features, the research team tested seven feature combinations listed in Table 2.

Table 2 Feature Input for Different Environmental Feature Combinations

Experiment number	1	2	3	4	5	6	7
Input	LST, FVC, and TVDI	LST and FVC	LST and TVDI	FVC and TVDI	LST	FVC	TVDI

Table 3 shows the results of different feature combinations in different various geospatial environments. The researchers evaluated the performance of different feature combinations in terms of accuracy, precision, and recall. Overall, the feature combinations of (LST, FVC, TVDI) and (LST, FVC) have the best performance in both rural and urban areas (Figure 13).

Table 3 Performance of Different Feature Combinations at Various Geospatial Environments

Features	Testing area	recall	accuracy	precision
LST, FVC, TVDI	Rural	0.87	0.854	0.89
	Urban	0.51	0.872	0.94
	Rural and Urban	0.86	0.857	0.86
LST	Rural	0.80	0.79	0.79
	Urban	0.56	0.86	0.66
	Rural and Urban	0.81	0.81	0.81
FVC	Rural	0.79	0.79	0.78
	Urban	0.50	0.5	0.43
	Rural and Urban	0.80	0.81	0.81
TVDI	Rural	0.50	0.42	0.21
	Urban	0.5	0.5	0.43
	Rural and Urban	0.5	0.5	0.25
LST, FVC	Rural	0.78	0.77	0.78
	Urban	0.86	0.85	0.85
	Rural and Urban	0.79	0.79	0.79
LST, TVDI	Rural	0.58	0.52	0.58
	Urban	0.5	0.23	0.33
	Rural and Urban	0.5	`0.52	0.5
FVC, TVDI	Rural	0.58	0.52	0.34
	Urban	0.5	0.27	0.58
	Rural and Urban	0.5	0.36	0.5

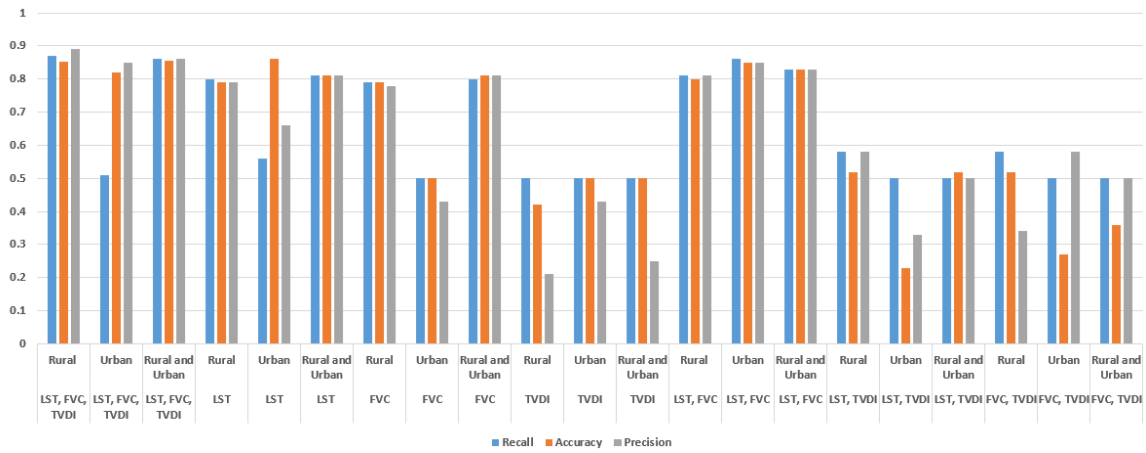


Figure 11 Performance of Seven Feature Combinations on Different Landcover

From the test results in Table 3, the research team conducted a comprehensive analysis and came to the following findings:

1) As Figure 11 shows, using a combination of multiple features as input outperforms using a single feature as the input. Overall, the feature combinations of (LST, FVC, TVDI), and (LST, FVC) have the best performance.

2) All the feature combinations, except for using TVDI, have a more reliable performance in rural areas than urban areas. The feature combination of TVDI has unreliable performance in both rural and urban areas. The performance of TVDI in rural areas is worse than the performance of TVDI in urban areas (Figure 11).

3) For Figure 12, the importance of a single feature has a sequence as LST>FVC> TVDI from strong to weak. LST is the most reliable environmental feature to detect canal leakage while TVDI has the worst performance.

4) The research team also analyzed how adding environmental features may influence the performance of existing feature combinations. From Figure 13-15, the research team compared how added environmental features can affect the current feature combination. The results showed that adding LST and FVC can improve the performance of the existing feature combination. Whereas, adding TVDI tends to damage the performance of the algorithm. The potential reason for this phenomenon is that the algorithm using Landsat 8, a medium resolution satellite image, to estimate soil humidity is not reliable.

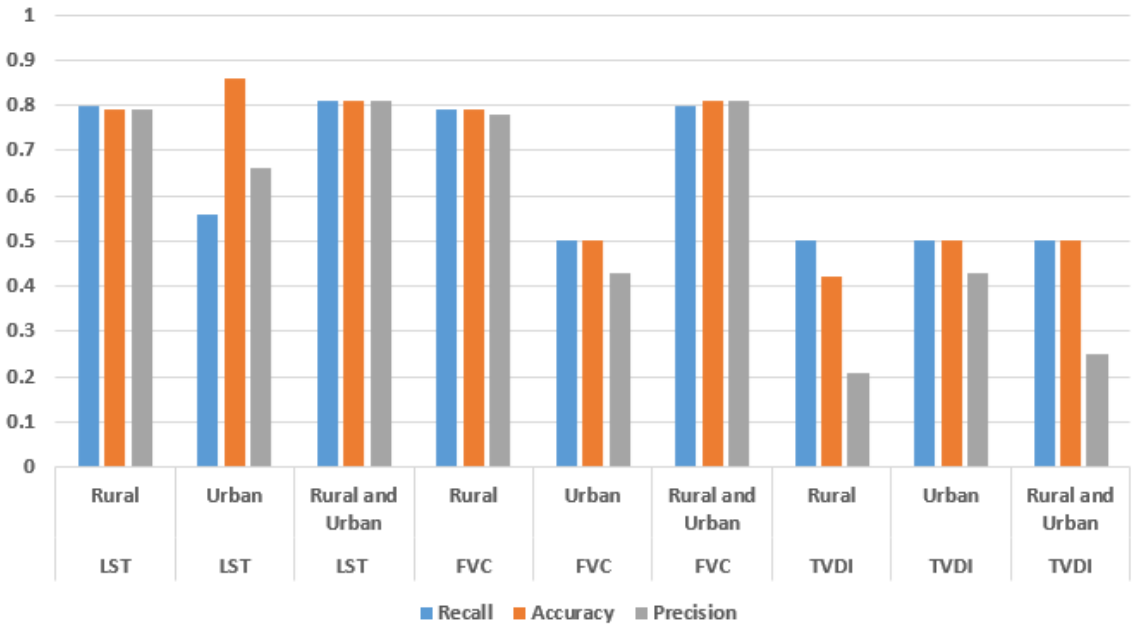


Figure 12 Performance of Single Environmental Feature on Different Landcover

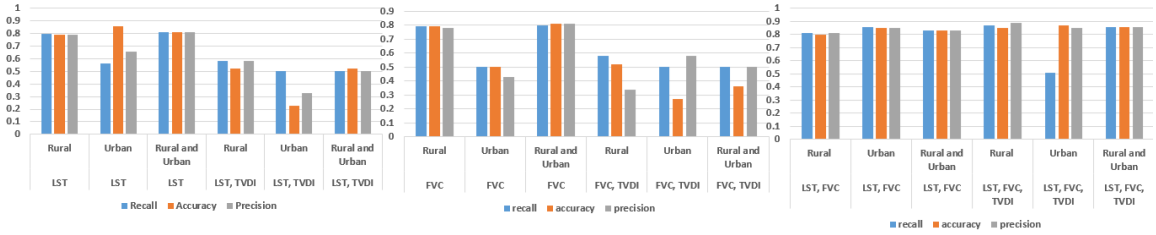


Figure 13 Performance of the Proposed Algorithm after Adding TVDI to Existing Feature Combinations

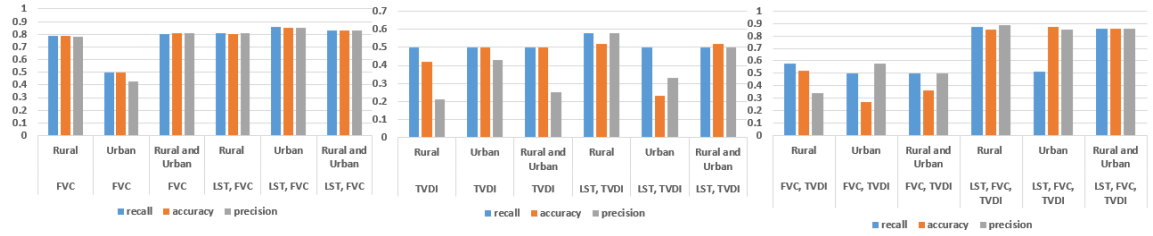


Figure 14 Performance of the Proposed Algorithm after Adding LST to Existing Feature Combinations

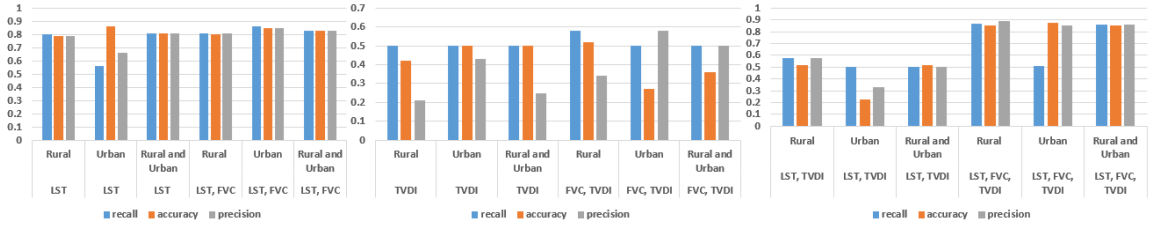


Figure 15 Performance of the Proposed Algorithm after Adding FVC to Existing Feature Combinations.

3.5.4 Comparison between the developed PGNN approach versus a conventional deep learning algorithm trained on raw satellite images

The authors developed the PGNN approach to increase the interpretation of the neural network. The expectation is that PGNN could help engineers explain and understand how environmental conditions could indicate the canal leaking conditions. To fully explore the value of the PGNN approach, the authors compared the developed PGNN approach against the conventional deep learning algorithms trained on raw satellite images. The authors used the same augmented dataset with 8,000 images as training and 2,000 images as testing. The only difference is that for the conventional deep learning approach, the authors used raw satellite images for training and testing the machine learning model.

For the PGNN approach, the authors used all three features, including LST, FVC, and TVDI. The authors used a desktop computer composed of an Intel CPU and an Nvidia 1080Ti graphics card. As for training time, the PGNN approach took 2 hours to train 10,000 iterations. The conventional deep learning approach took approximately 5 hours to train.

As shown in Figure 16, conventional deep learning can achieve similar accuracy, precision, and recall as the PGNN approach, although training the conventional deep learning model takes more time. Another major advantage of the PGNN approach is that the channels of input have physical meaning, which can help engineers understand which features are more useful for detecting canal leakages.

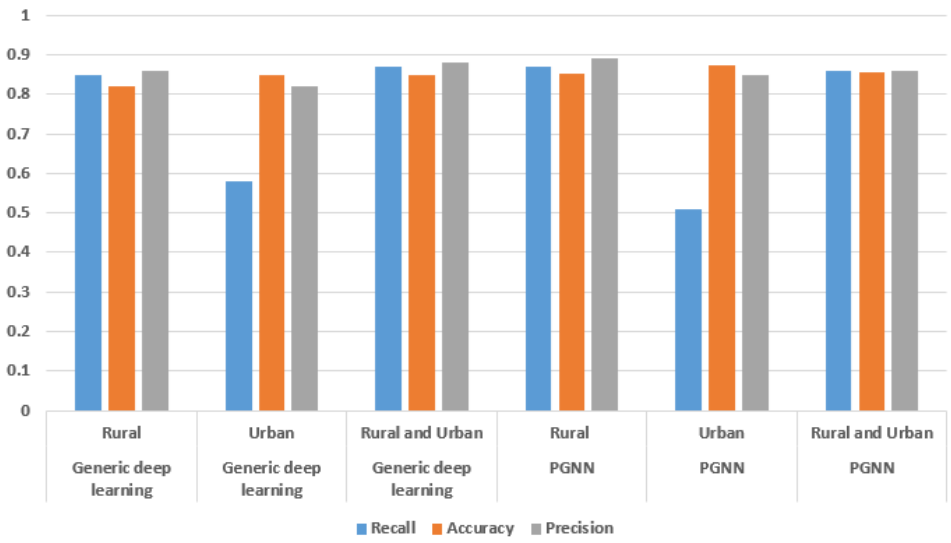


Figure 16 Comparison between a Conventional Deep Learning and the Developed PGNN Approach

3.6 Discussions

3.6.1 Implications

This section describes the potential implications of the developed PGNN anomaly detection to engineering practices. The neural network model built in this research will help engineers develop a decision-making process that relies on timely and high-quality field data and predictions based on such data. These automated remote sensory data processing techniques can help engineers to fully utilize large amounts of data available to them through multiple sources and expose them to large amounts of free remote sensory data sources, such as Landsat 8, Planet, and other satellite images potentially useful for civil infrastructure maintenance. These automated data processing algorithms can significantly reduce the data processing time and workforce needs through automated algorithms. Moreover, such a PGNN approach could help engineers to visualize the LST,

FVC, and TVDI in the targeted canal sections. The engineers could analyze the environmental features around the canal section and support the maintenance planning of canal sections.

3.6.2 Limitations

First, the limitation of this research is that the current algorithm can only classify canal sections into leaking and non-leaking categories because collecting and labeling the satellite images takes lots of effort and time. A more efficient method that can mark the satellite image into multiple levels of leakage will make the proposed algorithm more reliable and useful. Second, the resolution of the prediction results of the proposed algorithm is in the window-level. Each window is 8 pixels * 8 pixels, which covers a 240 meters * 240 meters area. To detect canal leakage with higher accuracy, the researchers may need to explore the use of the high-resolution satellite image such as Planet.

3.6.3 Future works

As for future work, the research team will improve the current work with the following parts to create more scientific and economic values. The first part will be extending the algorithm for additional anomaly detection tasks. The current algorithm can only classify canals as leaking and non-leaking because the established dataset has only two categories of labels. However, this algorithm could classify canal sections into multiple levels of water leakage if provided a data library that labeled canal sections into various levels of water leakage. The second part will be transplanting this method to other civil facilities such as underground pipes. The researchers will focus on producing a predictive spatiotemporal model that takes environmental conditions and any other physical and

technical parameters of civil infrastructure systems as “contextual conditions” to predict anomaly in civil infrastructures.

3.7 Conclusion

This research study explored the methodology of using multispectral satellite imagery to assist canal condition assessment. The research team used remote sensing algorithms to extract environmental features from multi-band satellite images. Then the research team adopted a PGNN that can automatically detect the leaking sections of canals on satellite images. To establish the dataset for the PGNN, the research team collected Landsat 8 satellite image and canal maintenance records from 2016 to 2019 in different areas.

Through the training and testing process, the proposed algorithm achieved the precision at 86%, recall at 86%, and accuracy at 85%. Furthermore, the research team tested different combinations of environmental features and explored how different combinations of environmental features influenced the performance of the developed algorithm in different geospatial environments.

The results indicate that a medium resolution satellite image can achieve excellent performance in canal leakage detection. A PGNN can achieve reliable classification accuracy, precision, and recall while being more computation efficient than conventional deep learning. Furthermore, the researchers found that the proposed algorithm has a better performance in rural areas than urban areas because the geospatial contexts in urban areas are more complicated than that of rural areas. To increase the interpretability of the neural network, the researchers integrated remote sensing knowledge to derive environmental features and tested how different combinations can influence the

performance of the proposed algorithm. The researchers found that LST tends to be the most important environmental feature among LST, FVC, and SH. Also, the researchers found that combinations of LST, FVC, and LST, FVC, and SH have the best performance in different landcover. These findings can provide an in-depth understanding of the interactions between environmental conditions and civil infrastructures. Moreover, these findings will lead to new knowledge about how environmental conditions influence civil infrastructure maintenance.

4 DETECTING ANOMALOUS WORKFLOW USING MULTIPLE OBJECT TRACKING WITH THE INTEGRATION OF CONTEXTUAL INFORMATION

4.1 Introduction

In recent years, with the emergence of affordable video cameras and advance of computer vision techniques, an increasing number of construction companies began to set up cameras on construction sites for field surveillance. Because most construction sites involve collaborative work, there are lots of interactions and communication between workers, workers, and machines (Yang et al. 2010). Tracking multiple workers is thus essential to supply information to analyze how different objects interact with each other. Multiple object tracking is a computer vision technology to locate multiple objects, maintaining their identities, and generate trajectories of different objects given an input video (Luo et al. 2014). As mentioned in (Luo et al. 2018), inaccurate detection and frequent identity switching are still the major problems of multiple object tracking. Previous studies did not systematically consider the identity switch in multiple object tracking, which can produce erroneous information from objects missed, mislabeled, or having discontinuities in the tracking process.

The current method adopted by the construction industry has inspectors for site observations and inspection to recognize unsafety behaviors and monitor the construction progress. Manual monitoring is time-consuming and error-prone and is not suitable for monitoring large construction sites with thousands of parallel activities (Zhu et al. 2017). Many researchers explored the potential of visual tracking to provide automated and

continuous monitoring (Cheng et al. 2013). Various challenges, such as occlusions and identity switch, have been bringing uncertainties into the tracking results, and no existing studies systematically examine and quantify such uncertainties. A systematic classification and synthesis of failures of multiple object tracking methods and relevant factors that cause those failures are thus crucial for quantifying decision risks based on information derived by multiple object tracking algorithms from field videos.

Waiting lines or queues exist in almost all industrial processes as well as construction sites (Akhavian and Behzadan 2014). The research team chooses a waiting time calculation of a queuing system as the testing application of multiple object tracking. During nuclear power plant outages, workers need to transport from the valve to the Radiation Protection Island (RPI, the space connecting the containment and the outside). Monitoring the waiting time workers spend RPI is important since delays in RPI could compromise the productivity and safety of the entire workflow (Zhang et al. 2017). The researchers proposed a multiple object tracking algorithm that uses video input to automatically the time different workers spend in the critical areas that related to the tasks with co-prerequisite or resource sharing relationships. Evaluation of the performance of multiple object tracking, in this case, is vital to assess the reliability of the waiting time calculation. In this research, the researchers reviewed the previous work of multiple object tracking. We proposed a multiple-object tracking algorithm to identify the scenarios where the algorithm calculated the waiting time with low precision and recall. The researchers summarized the failure scenarios to characterize the performance of multiple object tracking using the test case.

Furthermore, the researchers found that there are still challenges for reliable multiple object tracking. The researchers proposed to integrate context information to multiple worker tracking. Taking progress and quality monitoring, for instance, lots of methods developed for progress and quality monitoring often leverage strong priors such as a 3D Building Information Model (BIM) and schedule, which provide detailed information about the geometry and appearance of the elements in any point in time. Current methods used for tracking and analyzing activities of the equipment and labor does not use strong priors from BIM or schedule. Another typical example is proximity analysis using video data. Current MOT results are all calculated in pixel-level, which may hinder the usage of MOT in construction applications. Including contextual information, e.g., an operational layout map, can convert the MOT results to the metric unit, which makes the proximity calculation more applicable. Some researchers proposed to integrate context information to support the video reasoning process theoretically. However, the process of incorporating context information is a manual process by marking an area of interest in the image surface (Gong & Caldas, 2010). This marking process is inaccurate since the image from a single camera used to suffer from distortion and loss of depth problems (Gong & Caldas, 2010).

On the other hand, in the domain of computer science, integrating priors from context information could improve the performance of tracking with a single camera. Rather than tracking random people, an increasing number of researchers proposed to incorporate the information from contexts to guide more MOT algorithms for achieving more reliable tracking of multiple workers. Most surveillance cameras on construction sites are single

camera. Though some researchers tried stereo cameras to conduct 3D tracking to solve the loss of depth, employing a stereo camera for the large-scale construction sites can be impractical (H. S. Park et al., 2013). The region which can be monitored by a stereo camera is quite limited because only the overlapped areas of the two field of view of the dual cameras can be useful for 3D tracking.

Using MOT with a single camera for supporting high-level analysis needs rich context information and reliable MOT results. However, current methods cannot provide sufficient and accurate context information, and MOT suffers from loss of depth problem when using a single camera. Previous research did not develop a reliable method to register video and as-designed models for MOT. As-designed models, such as site layout maps, work zones, scene models, can provide extensive contextual information, which is essential for further analysis (Gong & Caldas, 2010). Tracking multiple workers using a single camera image is prone to inconsistent displacements. A consistent tracking algorithm must be able to track a worker regardless of his position in an environment. Consider the case when a worker approaches a single fixed camera. As the worker gets closer to the camera, the worker's displacement in the image space becomes larger and larger. In other words, the worker's velocity changes, although in the object space, the worker has a constant velocity. Now, consider another worker who moves away from the same camera. The worker's displacement becomes smaller and smaller, resulting in a lower speed in the image space. There could be other workers walking across the room, running, and standing still. The loss of depth caused the issues because construction sites have only a single fixed camera.

The innovative computer vision approach presented in this paper augments state-of-the-art multiple object tracking algorithms with contextual information from as-designed models, site layout drafts, and other relevant project data (e.g., schedules, as-designed model). The augmented algorithm can overcome the challenges of using one camera for reliable workspace surveillance in cluttered environments. The researchers performed object detection, which is the first step of MOT. Through a homograph transformation, the researchers register the video to the as-designed model (section 4.3.2). The researchers used the registration model to project detection results from image space to as-designed model space, which can mitigate the loss of depth problem. Finally, the author performs MOT in the as-designed model space.

4.2 Previous Research

Multiple object tracking (MOT) has gained lots of research interests in recent years due to its academic and commercial potentials (W. Luo et al., 2014). The information of objects generated from multiple object tracking can support further behavior analysis and action recognition. Due to the construction sites are complex and dynamic, involving lots of workers and equipment, the algorithm proposed by the researchers usually came across various problems such as missing objects and losses of tracks. Previous researchers evaluated the multiple object tracking regarding precision and recall (Xiao et al. 2018; Yang et al. 2010). Few of them gave a comprehensive review of where multiple object tracking fails in the field. This research tested the multiple object tracking algorithm in fourteen video clips (over 5,000 frames) to identify the scenarios where multiple object tracking fails. Previous studies described that identify switch remains as the main

challenge that hinders the practical application of multiple object tracking on construction sites (Xiaochun Luo, Li, Cao, Dai, et al., 2018). The author of this dissertation proposed to use contextual information to reduce the identity switch.

4.3 Research Methodology

To overcome the challenges of using one camera for reliable workspace surveillance in cluttered environments, the researcher transformed the detection results from the camera's image space into the floor of the RPI. With the projected detection results, the researcher associate detection results in different frames to get tracking results. The idea is to utilize a homograph transformation to track workers on the horizontal plane of the ground. In this section, the researchers presented the object detection algorithms for detecting workers in the video. Then the researchers explained how to integrate contextual information for multiple object tracking on construction sites. The researchers chose the Kalman filter and the Hungarian algorithm for tracking and data association.

4.3.1 Object detection

Recently, many researchers used deep neural networks (DNN) for object detection due to superior performance (D. Kim et al., 2019). The traditional object detection relies on features extracted by image descriptors. Extraction and selection of image descriptors can significantly influence object detection results. Moreover, conventional object performance detection methods suffer from different challenges, such as scale variation, occlusion, and viewpoint change (Golabchi et al., 2018). The deep neural networks can extract fine-grained features through training and testing processes, which enables more reliable object detection. Fed with the extensive training dataset, DNN can ensure the

robust and reliable performance confronted with different challenges, including occlusion, scale variation, and viewpoint change.

In the construction domain, many researchers have also explored the usage of DNN-based object detection for localization and tracking of construction entities. The authors listed recent object detection methods used in the construction domain in Table 4. Moreover, the authors attached the performance of these detectors tested from benchmark data (Redmon & Farhadi, n.d.). In Table 4, the mAP stands for mean Average Precision, and FPS stands for frame per second. mAP denotes the classification performance of the detector on different categories of objects. FPS represents the speed of the detector.

Table 4 DNN-based object detection used in construction domain (results of mAP and FPS are from **(Redmon & Farhadi, n.d.)**)

Model	Related studies	mAP	FPS (speed)
YOLO-V3	Detect non-hardhat-use (Fang et al., 2018) Detect construction entities for proximity monitoring (D. Kim et al., 2019)	55.3%	35
Faster R-CNN	Detect construction-related objects for matching activity patterns (Xiaochun Luo, Li, Cao, Dai, et al., 2018)	42.7%	17
R-FCN	Detect equipment in tunnel earthmoving process (H. Kim et al., 2018)	51.9%	12
SSD	Detect cranes for monitoring safety hazards (Roberts et al., 2017)	45.4%	16

Among these different detection methods, the researchers chose YOLO-v3 for this study. YOLO-v3 has proven to be one of the state-of-the-art object detection approaches due to accuracy and speed.

4.3.2 Homograph transformation for the integration of contextual information

Homograph transformation is a widely used geo-referencing technique that can transform one coordinate system to another coordinate system. Homogenous points are the pairs of points that present the same locations in different coordinate systems. Given several pairs of homogeneous coordinates, the algorithm can calculate the transition matrix between two coordinate systems.

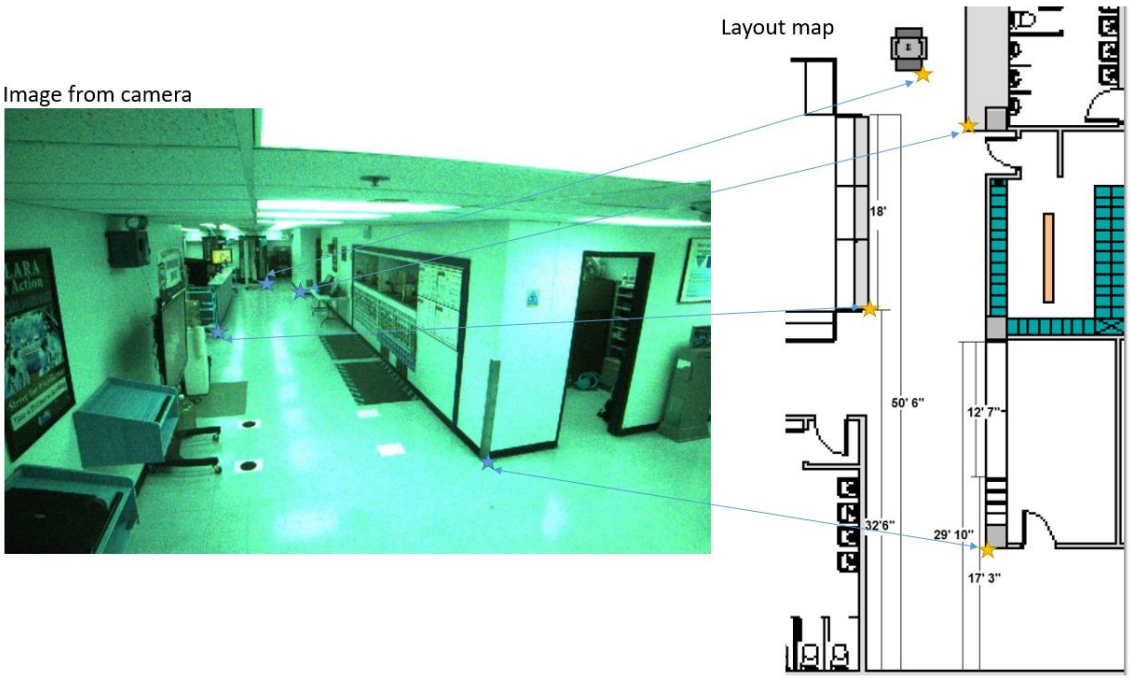


Figure 17 Select Homogeneous Points from the Image of the Camera and Layout Map. Stars Denote the Selected Homogenous Points.

Because construction sites are usually crowded and changing rapidly, using a camera for multi-object tracking suffers from occlusion and loss of depth problem described in (D. Kim et al., 2019). Loss of depth problem can lead to an inaccurate estimation of objects' location and thereby affect the performance of multi-object tracking. On construction sites, design maps or layout maps are usually accessible that can provide accurate contextual information on the geometry of the construction sites. In this research, the authors proposed to integrate the geometric contextual information to improve the multi-object tracking performance.

As seen in Figure 17, the authors selected four pairs of homogenous points marked by a star. The selection strategy is to make the point spread the whole map, and all the points should not be collinear.

The authors used (x, y) to denote the point coordinates in the image captured by a camera and (X, Y) to present the point coordinates in the design map coordinate system. Then the authors build the transformation matrix as below:

$$\begin{vmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -y_1X_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2X_2 & -y_2X_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3X_3 & -y_3X_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4X_4 & -y_4X_4 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y_1 & -y_1Y_1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2Y_2 & -y_2Y_2 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3Y_3 & -y_3Y_3 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4Y_4 & -y_4Y_4 \end{vmatrix} \cdot H = \begin{vmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{vmatrix} \quad \text{Equation 9}$$

Where H is an eight-dimensional vector:

$$H = |h_1 \quad h_2 \quad h_3 \quad h_4 \quad h_5 \quad h_6 \quad h_7 \quad h_8|^T \quad \text{Equation 10}$$

Given the values of $(x_{1\sim4}, y_{1\sim4})$ and $(X_{1\sim4}, Y_{1\sim4})$, the authors can get the transformation matrix. For a new pair of coordinates from the camera image, the authors can calculate the corresponding coordinates in the design map using the following formula:

$$\begin{cases} X = \frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + 1} \\ Y = \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + 1} \end{cases} \quad \text{Equation 11}$$

4.3.3 Multiple object tracking

Multiple object tracking is a computer vision technology to locate and track multiple objects (X. Li et al., 2010). As mentioned in (Xiaochun Luo, Li, Cao, Dai, et al., 2018), inaccurate detection and frequent identity switching remain as the significant challenges that hinder the practical usage of multiple object tracking in construction management. The following section describes the essentials of multiple object tracking using Kalman filter and Hungarian data association.

After getting the detection results and projecting the detection results on to layout map, the next step is to use the motion model to estimate the object's location in the coming frame given the current location. A typical and widely used model is Kalman filter (Andriyenko & Schindler, 2011), and the authors gave a detailed description in Section 4.3.3.1. Then for the objects in the new frame, the engineers will have two sets of locations of the objects. The motion model calculates one set of locations. The object detection algorithm estimates the other set of locations. A data association algorithm needs to assign detections with existing objects (Bewley et al., 2016). Section 4.3.3.2 introduces a Hungarian data association method.

4.3.3.1 Kalman filter for multi-object tracking

Kalman filter has been popular in many domains, such as mechanical engineering and robotics, to predict and model the states of linear processes (X. Li et al., 2010). Kalman filter is efficient and practical for building a multi-object tracking. To use the Kalman filter to estimate object state, the authors use x_t to present the state of object at time t, which is a four-dimensional state vector $x_t = [x_{c,t},\ y_{c,t}, v_{x,t}, v_{y,t}]$.

4.3.3.2 Hungarian data association

The Hungarian algorithm can associate objects across frames by minimizing the designed cost functions (Bewley et al., 2016). The authors chose to use the Hungarian algorithm to associate objects from one frame to another. To describe the distance between objects, the author used the distance function (X. Li et al., 2010) to measure the distance between i^{th} object in t frame and the j^{th} in $t + 1$ frame.

4.3.4 Evaluation metrics

To evaluate the multi-object tracking performance, the authors reviewed previous relevant research (M.-W. Park & Brilakis, 2016) in construction. Previous studies use recall, precision, and accuracy to assess multiple worker tracking algorithms (M. W. Park & Brilakis, 2012). However, these metrics cannot assess the performance of the algorithm for tracking the same person. As mentioned in (Xiaochun Luo, Li, Cao, Dai, et al., 2018), the identity switch is a challenge that can make the tracking performance unreliable. In crowded scenarios,

- FP: If the predicted location of the worker does not match the ground truth, the false prediction is FP, which is also called a false alarm.
- FN: If the algorithm did not track the worker location, the unassigned ground truth location is a false negative.
- ID_switch: an ID_switch happened when the identity numbers of two objects switch from the previous frame to the current frame.
- FM: FM denotes track fragmentations. FM counts how many times a ground truth trajectory loses track.
- MOTA (Multiple Object Tracking Accuracy): MOTA is a widely used metric to evaluate multi-object tracking performance.
- MOTP (Multiple Object Tracking Precision):

$$Precision = TP/(TP + FP) \quad \text{Equation 12}$$

$$Recall = TP/(TP + FN) \quad \text{Equation 13}$$

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad \text{Equation 14}$$

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + ID\_switch_t)}{\sum_t gt_t} \quad \text{Equation 15}$$

$$MOTP = \frac{\sum_{t,i} d_{t,i}}{\sum_t C_t} \quad \text{Equation 16}$$

In the equations above:

- c_t is the number of matched predicted objects in frame t.

- $d_{t,i}$ is the overlapped bounding box area between predicted object i and the ground truth object.

4.4 Experiments

The researchers put a camera on the Radiation Protection Island (RPI) the outage in a Nuclear Power Plant and collected 24-hour video data. The researcher used a laptop that was placed in the RPI and not needs to be connected to the network and will not be live streaming. The researchers selected seven video clips to test the algorithm. Also, the researchers subsampled all the chosen to test if the performance will be affected when lowering the video resolution. In total, the researchers have 14 video clips to evaluate the performance of the algorithm. For each video clip, the researchers tested the algorithms for every worker showed up in that clip. The researchers used the evaluation metrics to compare how the test integration can affect the performance of the multiple object tracking on construction.

4.5 Results

4.5.1 Comparison of Tracking on layout map and tracking in video space

Previous research considered identity switch as one major challenge that hinders the practical application of multiple object tracking on construction sites (Luo et al., 2018; Seo et al., 2015). To validate the effectiveness of the developed algorithm, the author of

this dissertation compared how the integration of contextual information can affect the performance of object tracking on construction sites.

Table 5 Comparison of Tracking with or without Contextual Integration (up Arrow Denotes that the Metric is the Larger, the Better, Vice Versa)

	Recall ↑	**Precision** ↑	**FP** ↓	**FN** ↓	**ID_switch** ↓	**FM** ↓	**MOTA** ↑	**MOTP** ↑
Without contextual integration	41.1	56.7	592	1,166	32	98	27.2	57.3
With contextual integration	43.2	58.2	576	1,052	24	91	30.7	63.8

The author selected eight video clips and tested the developed algorithm. Table 5 shows the performance of the developed algorithms with and without the integration of contextual information. After the integration of contextual information, ID_switch has reduced 25% and FP, FN also decreased.

4.5.2 Performance of tracking workflow

This section described the performance of the algorithm of tracking multiple workers. From the video clips in which the algorithm calculated the waiting time with low recall and precision, the researchers investigated those videos and analyzed the scenarios where the algorithm fails. This study investigated the performance of the proposed tracking algorithm in over fifty hours long videos collected from an indoor space to quantify how reliable the computer vision algorithm could calculate the average waiting times of workers in queues at different areas of a workspace. The researchers reviewed all the

videos and selected fourteen video clips that are of the lengths of 30 to 60 seconds. Each video contains 300 to 400 frames. These fourteen videos include challenging scenarios with severe occlusions, scale variations. The researchers tested the developed algorithm on the selected video clips. The authors characterized the impacts of four factors, including occlusions, number of people in the video, temporal resolution, and spatial resolution on the precision and recall of estimation of waiting time produced by the visual tracking algorithms.

The authors labeled the ground truth of the objects of interest manually in the tested videos. Then the tracking results from the proposed method were compared with the ground truths to calculate the precision and recall. The Euclidean distance between center locations of the objects and their corresponding ground truths in the videos measures the tracking precision. The unit of the distance is pixel (Zhu et al., 2017). However, as for the practical construction application in this research, the research team chose to evaluate the tracking performance in terms of accuracy and precision of waiting time estimation. The reason is that the final objective of this multiple object tracking algorithm is to monitor and predict the waiting time of workers.

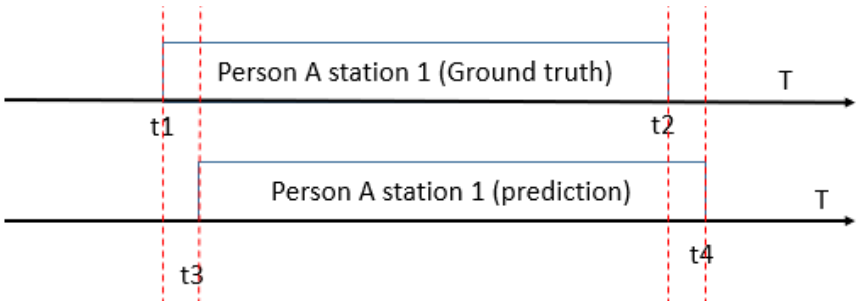


Figure 18 Example of Performance Evaluation

$$Recall = \frac{t2-t3}{t2-t1} \qquad Precision = \frac{t2-t3}{t4-t3}$$ Equation 17

Table 6 Test Results Characterizing the Number of Workers, Occlusion Level, Time Resolution, Spatial Resolution (Numbers Highlighted by Red and Bold Font Indicate Low Recall and Precision)

ID	**number of workers**	**Occlusion level**	**Time resolution**	**Spatial resolution**	**Average Precision**	**Average Recall**
1	4-6	High	6	968*608	0.98	0.77
2	2-3	no	6	968*608	0.97	0.54
3	1-3	medium	6	968*608	1	0.99
4	1	no	6	968*608	**0.32**	**0.1**
5	1-3	no	6	968*608	1	**0.15**
6	1-3	no	6	968*608	0.70	0.58
7	1	no	6	968*608	1	0.61
8	4-6	High	6	600*600	0.5	0.17
9	2-3	no	6	600*600	**0**	**0**
10	1-3	medium	6	600*600	1	**0.05**
11	1	no	6	600*600	**0.1**	**0.05**
12	1-3	no	6	600*600	0.87	0.1
13	1-3	no	6	600*600	**0.43**	**0.32**
14	1	no	6	600*600	1	0.94
Average					0.70	0.38

As Figure 18 and Equation 17 show, the green area in the time axis represents the actual value of the time a person A stayed in station one, and the blue line gives the value predicted by the computer vision algorithm. The researchers calculated the recall and precision of the proposed algorithm. The recall means the percentage of the time predicted correctly by the computer vision algorithm among the real-time duration. The precision means among the duration of prediction, the percentage of the correct prediction.

4.5.3 Scenarios where multiple object tracking failed

As Table 6 shows, the algorithm can achieve excellent precision on the collected data, the average precision of the tested 14 videos is 0.70. The average precision means that the algorithm can calculate the waiting time of workers in stations at a 70% level. The average recall of the tested videos is 0.38. The average recall means for the time the workers spent in stations, the algorithm can track 38% of the time.

Figure 19 ID Switch due to Inter-Worker Occlusion

Figure 20 False Detection due to Reflective Objects (Red Circle the False Detection, the Algorithm Detected One Worker in the Video, whereas there is No Worker)

From the tested results, the researchers identified four typical scenarios where the algorithms were likely to fail based on the data available. The researchers will plan to collect more data for more comprehensive testing of the developed algorithms in additional scenarios. The first scenario is that when irrelevant workers passed the station and occluded

the workers. As Figure 19 shows, the algorithm assigned id 2 in the left image to the worker in the station. When the workers passed the station, the id 2 went to another person in the middle picture. In the right image, id two missed. This scenario is an example of an identity switch, which causes the calculation of waiting time inaccurate.

Another typical failure is false detection in Figure 20. Sometimes the algorithms could detect more people than the number of workers in the video. Due to reflections of the mirror in the RPI room, the current state-of-art algorithm could give false detection of the worker, which also makes the waiting time inaccurate. A similar problem could happen when there are reflective objects on the construction sites.

Figure 21 Occlusions due to the Background Obstacles. (The Algorithm Missed the Worker at the Left.)

The research team found the algorithm calculated the waiting time with low recall and precision when the background obstacles occluded the worker. Figure 21 shows that the algorithm missed the worker at the left of the scene because the wall occluded part of his body. Occlusion by the background obstacle happened a lot in real construction sites such as occlusion from excavator and wall.

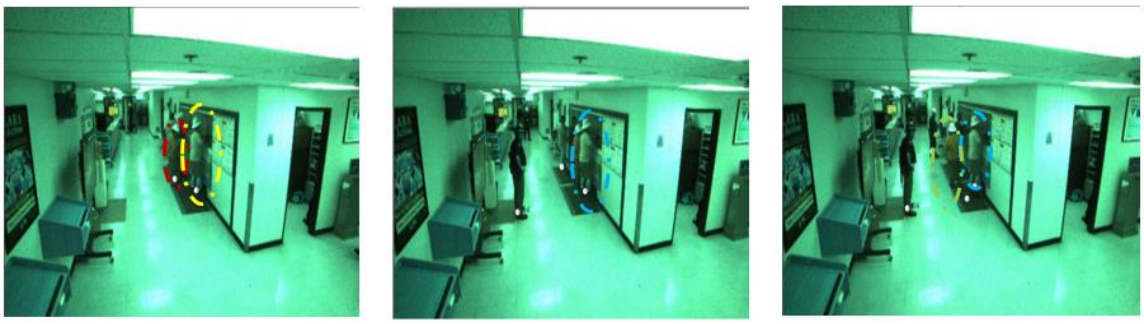


Figure 22 Missed Objects due to Workers Merge and Split

Another typical failure is that the algorithm missed the worker when they merge and split. As shown in Figure 22, in the left image, the two workers circled in red and yellow merge together at first. Then in the middle picture, the algorithm considered them as a new object together. In the right image, when they split, the algorithm assigned a new identity to the two workers, which means the algorithm failed to track the worker continuously and assigned a new identity to the worker.

4.6 Discussions

4.6.1 Implications

This section describes the potential implications of the developed anomalous workflow detection approach to engineering practices. The developed computer vision-based multiple object tracking algorithm shows the feasibility of using a static camera for construction project workflow monitoring. The integration of contextual information can improve the performance of multiple object tracking algorithm. For the construction domain, this camera-based tracking approach could mitigate the privacy concern of workers. Different from GPS, the developed approach does not need to install any device on workers' bodies. For the computer science domain, this research shows that

integrating contextual information for construction applications could improve the performance of the algorithms.

4.6.2 Limitations

Tracking multiple objects on dynamic and complex construction sites is still a challenging task. According to previous research (Redmon & Farhadi, n.d.; Shiv Kumar et al., 2019; Zhou & Tuzel, 2017), object detection algorithms can achieve high accuracy and precision. However, continuously tracking the same object without losing track remains difficult. This research developed a method of integrating contextual information to improve the performance of multiple object tracking. There are still mainly the following factors which bring difficulties to multiple object tracking on construction sites:

1) Occlusion problem: occlusion frequently occurs on construction sites. Workers may occlude with each other when they collaborate and stay close to each other. Workers may occlude with equipment and materials on construction sites such as scaffolding and lorry(Xiaochun Luo, Li, Cao, Dai, et al., 2018).
2) Appearance: workers on construction sites need to follow safety rules by wearing uniform hardhat and protection equipment. Admittedly, these policies can make workers' appearance more uniform and assist the safety management such as non-hardhat-use detection (Fang et al., 2018; M.-W. Park et al., 2015). Nevertheless, as for multiple object tracking, this uniform appearance will increase the difficulties of the matching process in the multiple object tracking.

3) Complicated activities and pose change involved: different from pedestrian tracking, which many computer science researchers focused on (Milan et al., 2016; Robicquet et al., 2016). Construction workers need to conduct diverse activities to finish the project. The same worker needs to have a different posture when performing the tasks. The pose change of the same worker may bring difficulties to continuous tracking.

4.6.3 Future research direction

The proposed approach for improving multiple worker tracking aims to integrate domain knowledge of construction into multiple object tracking. Currently, the developed contextual information can provide accurate geometric details of construction sites, thereby increasing the reliability of multiple object tracking. However, the authors will continue to develop a more generalized integration method that can integrate more construction information such as workflow, schedule, and worker role information into multiple object tracking.

Another research direction that the authors will explore to extend this multiple object tracking is to integrate contextual information into 3D object tracking. The developed algorithm in this research focused on 2D multiple object tracking. 3D tracking of construction sites is more complicated, considering the dynamic and complex environments on construction sites (Lee & Park, 2018).

4.7 Conclusion

This research presented an approach of integrating contextual information to improve multiple object tracking on construction sites. From the tested results, the researchers generalized four typical scenarios where multiple object tracking may fail in construction applications. Moreover, the tested results showed that integrating contextual information can reduce identity switch problem and increase the reliability of continuous tracking without losing objects' identities.

Based on findings in this systematic characterization of the multiple object tracking algorithms and the uncertainties of waiting time estimation, future research will be developing more extensible integration methods to use domain knowledge and information to improve the multiple object tracking results in construction applications.

5 DETECTING ANOMALOUS BEHAVIORS OF AIR TRAFFIC CONTROLLERS FROM TIME SERIES OF FACIAL EXPRESSIONS AND HEAD POSES

5.1 Introduction

Air Traffic Controllers (ATCs) provide essential information and instruction to pilots and allows the pilot to maintain safe separation distances between aircraft. Facial expressions and head pose of ATCs could be indicators of changes in the ATCs-pilot communication patterns and signify potential operational errors. For example, ATCs may miss a read-back failure of a pilot due to fatigue. The Federal Aviation Administration (FAA) has introduced well-designed Standard Operating Procedures (SOPs) to reduce the impacts of the anomalous behaviors of controllers (i.e., distraction, confusion) of aviation safety. However, human errors still exist and contribute to more than 70% of all aviation accidents in the United States (U.S.). Human behavioral monitoring of ATCs is thus necessary for effectively recognizing anomalous behaviors and human error prediction during an air traffic control (Liu & Goebel, 2018).

During air traffic control processes, ATCs need to monitor the movements of aircraft and instruct pilots through radio communications (Mosier et al., 2013). Most of the ATCs' tasks are cognitive tasks, which means the tasks have a high reliance on mental processes (Isaac et al., 2002). Besides, those tasks also require ATCs to move their head and pay attention to the moving aircraft. Computer vision techniques could provide clues for understanding the cognitive functions of ATCs during traffic control by extracting facial expressions, head poses, and eye blinks.

According to the National Transportation Safety Board (NTSB), researchers identified poor ATCs behaviors as one of the most significant signs that lead to human errors (Crutchfield, 2005; Nealley & Gawron, 2015). Wu also claimed that human errors caused 75% of the accidents, and fatigue related to 21% of the accidents(F. Wu et al., 2015). Sarter analyzed the NASA Aviation Safety Report System incident reports in terms of the formal characteristics of underlying errors, the cognitive stage, and the behavior level at which these errors occurred (Sarter, 2009). Most incidents involved lapses (i.e., failures to perform a required action) or mistakes. Ameen identified fatigue as one of the factors that directly affect human behavior in terms of accuracy and reaction time (Ameen, 2014). However, most of these errors were detected based on routine checks and the observed outcome of an action, respectively. Behaviors of ATCs are thus significant for ensuring the safety of the National Airspace System (NAS). An effective human behavior monitoring (e.g., fatigue detection, distraction detection, and so on) system is thus necessary to not only help detect anomalous human behaviors but also predict human errors and avoid ATCs-related accidents.

Facial behaviors are the primary source of information for fatigue and distraction detection (Ameen, 2014). Facial behaviors can be natural and important means for human beings to communicate their emotions and intentions (Shan et al., 2009). Automatic facial expression analysis has critical applications in many areas, such as human performance monitoring (Bailenson et al., 2008). Previous studies have examined the feasibility of using computer vision techniques for human behavior monitoring through facial expression analysis (Ameen, 2014; Shan et al., 2009). Soukupová has developed a real-

time eye blink detection algorithms using key points of human faces to detect human operators' vigilance (e.g., driver drowsiness) (Soukupová, 2016). Reddy has developed a real-time driver drowsiness detection system by using facial landmarks of the drivers (Reddy et al., 2017).

These studies show the potential of using facial behavior analysis for human workload and fatigue monitoring. However, few studies examined how time series of facial expressions and head pose of ATCs could capture the temporal patterns of ATCs' mental and physical states. This study developed a computer vision-based real-time facial-expression and head-pose analysis to detect anomalous behaviors of ATCs.

5.2 Previous Research

5.2.1 ATCs performance monitoring

Emotional states (e.g., happy, surprise, sad) have received a great deal of attention due to their impact on analyzing the cognitive status and mental and physical health (Truschzinski et al., 2018). Considering the significance of workers' emotions in construction activities, some researchers have attended to investigate the emotional states of construction workers. Although various psychological instruments have been widely used for monitoring emotional states quantitatively, most of the methodologies are survey-based, which may be influenced by subjective opinions. Besides, the survey-based method may have an impact on the current work of the construction workers' ongoing work (Bhandari et al., 2016). Therefore, more recent research attempted to monitor human emotions according to physiological responses. Researchers tried to investigate the relationship between work performance and the typical physiological reactions,

including heart rate (HR) and electroencephalogram (EEG). (Hwang et al., 2018) demonstrated the applicability of a wearable EEG sensor for measuring workers' emotions, particularly valence levels. (Jiayu Chen et al., 2017) introduces a novel EEG approach to estimate task mental workload in construction projects.

5.3 Research Methodology

This section first provides a brief description of the proposed computer vision-based model for the timing analysis of construction workers. Then, the author of this dissertation describes the technical details and algorithms involved in this system. The goal of this study is to develop a methodology that can identify anomalous facial behavior to help monitor ATCs' behaviors. The proposed research consists of three modules:

1) design and implement a human-in-the-loop simulation for collected ATCs operation data.

2) use computer vision algorithms to extract time-series data of facial expression, eye blink, and head pose, which can reflect the mental or physical state of ATCs.

3) develop an anomaly detection algorithm for multivariate time series data extracted from the computer vision algorithms. The final output of this algorithm is the anomalous timeslots when the ATCs' behaviors change (Figure 23).

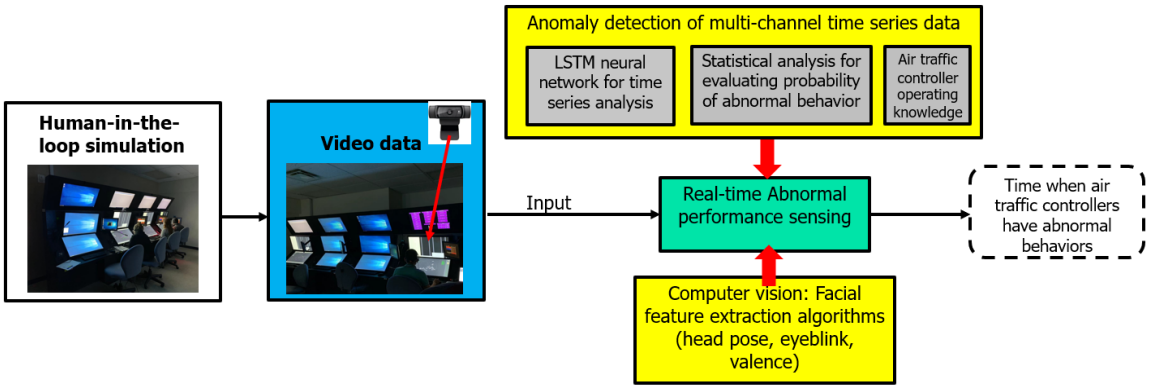


Figure 23 Overall Framework of the Abnormal Behaviors Detection of ATCs' Behaviors

5.3.1 Facial feature extraction

ATCs provide essential instructions to pilots and assist the pilot in maintaining safe separation distances between aircraft. The performance of ATCs is a critical factor in ensuring safe NAS operations. According to the previous studies, the authors found that attention and communication contributed more than 70% of the operational errors (Ligda et al., 2019). During air traffic control processes, ATCs need to monitor the trajectories of several aircraft and communicate with pilots through radio. Most of ATCs' operations are cognitive tasks that require a high workload on mental processes.

In this research, the authors used head pose, eye blink, and facial expressions as the three indicators of changes in the human performance of ATCs and to achieve an in-time warning of potential operational errors. The computer vision techniques could extract the head pose, eye blink, and facial expressions in a real-time way to assist in-time monitoring of ATCs' performance.

5.3.1.1 Head pose estimation

Head pose estimation refers to the ability to infer the orientation of a person's head relative to the view of the camera (Murphy-Chutorian & Trivedi, n.d.). Previous research introduced three degrees of freedom (DOF) to characterize the pose of the head by pitch, roll, and yaw angles (Figure 24)(Brolly et al., 2003; Murphy-Chutorian & Trivedi, n.d.; Ruiz et al., 2017). The research team implemented a real-time head pose estimation method, as pictured in Figure 24 (Amos et al., 2016). The technique developed by Tadas and his colleagues took an input image and conducted a face detection. Facial landmark refers to a set of points which can represent facial components and facial contour(Y. Wu & Ji, n.d.). The algorithm uses facial landmark detection to retrieve the key points to represent a human face. Then the algorithm estimated head pose using the facial landmark points.

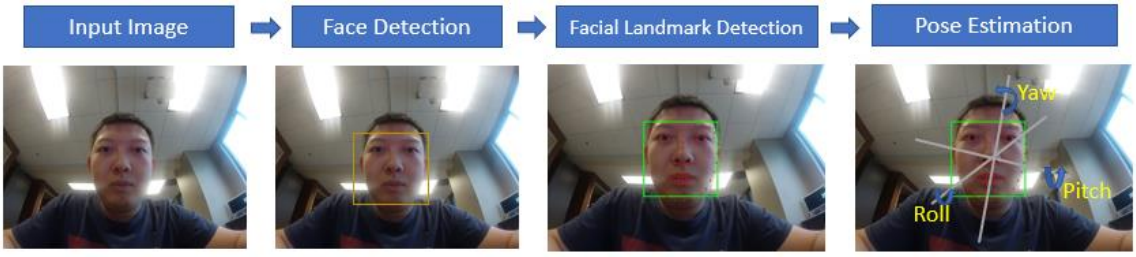


Figure 24 Real-Time Face Pose Estimation

ATCs usually operate with multiple computer displays. The head pose can be an indicator of ATCs' visual attention. Analysis of ATCs' head pose can provide information about the status of ATCs and identify abnormal behaviors of ATCs. With the emergence of practical computer vision algorithms and low-cost cameras, it is possible to achieve a real-time head pose estimation using a single camera. In the computer vision domain,

head pose estimation is the technique of estimating the orientation of a person's head using images regardless of identity. The authors adopted a markerless head pose estimation algorithm. This algorithm takes the input image and conducted face detection to identify facial landmarks. Facial landmarks are a set of 2D points that can represent facial components and facial contour (Y. Wu & Ji, n.d.). Meanwhile, a generic 3D model can provide the 3D coordinate of the corresponding 2D facial landmarks. Solving this perspective-n-point problem can obtain the results of head pose estimation. The perspective-n-point is to find the transformation matrix between a 3D coordinate system and a 2D coordinate system with known n pairs of 2D coordinates and 3D coordinates of the feature points. As

$$p = A[R|t]P \quad \text{Equation 18}$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_x \\ r_{10} & r_{11} & r_{12} & t_y \\ r_{20} & r_{21} & r_{22} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \text{Equation 19}$$

$$\begin{cases} yaw = atan2(r_{10}, r_{00}) \\ pitch = atan2(-r_{20}, sqrt(r_{21}^2 + r_{22}^2)) \\ roll = atan2(r_{21}, r_{22}) \end{cases} \quad \text{Equation 20}$$

In the above equation Equation 18 - Equation 20:

- x, y represent the 2D coordinates and X, Y, Z represent the 3D coordinate
- f_x, f_y, c_x, c_y are the parameters of the camera
- Yaw, pitch, roll represent the head pose angle.

5.3.1.2 Eye Blink Detection

Detecting eye blinks is essential in systems that monitor a human operator's vigilance, such as driver drowsiness (Danisman et al., 2010). Such an eye-blinking analysis system could warn ATCs staring at the screen without blinking for a long time to prevent erroneous operations due to dist. According to a previous study (Soukupová, 2016), real-time facial landmark detectors can capture most of the characteristic points on a human face image, including eye corners and eyelids. The researchers trained the facial landmark detector on a large and diverse dataset, which makes the detector robust to different illumination, various facial expressions. For every video frame, the eye aspect ratio (EAR) between the height and width of the eye is computed according to EAR=

$$\mathrm{EAR} = \frac{||p_2 - p_6|| + ||p_3 - p_5||}{2||p_1 - p_4||} \qquad \text{Equation 21}$$

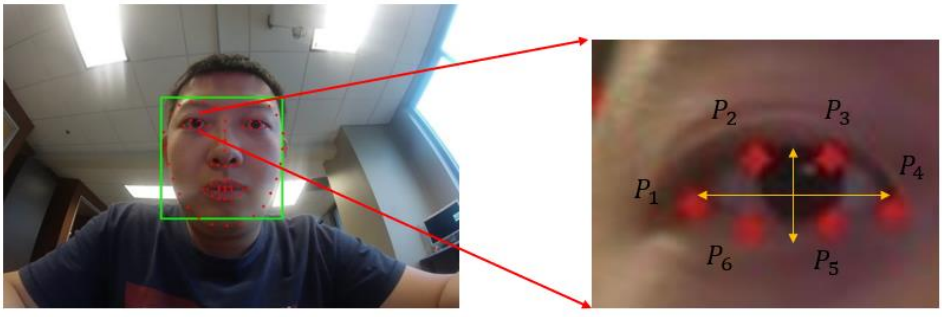


Figure 25 Eye Blinking Detection Using EAR

Equation 21. In this equation, $p_1, \ldots, p_6$ are the 2D

5.3.1.3 Facial expression classification

Usually, a person conveys emotion when reacting to a particular event. Emotions can be characterized as negative (sadness, anger, or fear), positive (happiness or surprise), or neutral. The earliest method for characterizing the physical expression of emotions is the Facial Action Coding System (FACS) (Paul Ekman, 1997). FACS defined a set of facial muscle movements and their corresponding emotions. In recent years, FER has attracted lots of researchers' attention due to its practical values in robotics, driver fatigue detection, and other human-computer interactions (S. Li & Deng, n.d.). The traditional methods, such as local binary patterns (LBP) (Shan et al., 2009), used handcrafted features to classify facial expression. More recently, with the development of deep learning and Graphic Processing Units, FER has transferred to deep learning methods, which achieved state-of-the-art recognition accuracy. Provided more diverse and labeled training data, deep learning-based FER show advantages over the traditional methods.

Arriaga and his colleagues proposed a real-time convolutional neural network (CNN) to achieve facial expression classification (Arriaga et al., n.d.). The research team used this model because this model can have excellent performance on hardware-constrained systems, and the speed can achieve real-time performance. The researcher used the FER-2013 dataset to implement the model in Figure 26. The FER-2013 dataset contains 35,887 grayscale images where each image belongs to one of the following classes: "angry," "disgust," "fear," "happy," "sad," "surprise," "neutral"}. The research team used the trained neural network to recognize the emotion of real-time video.

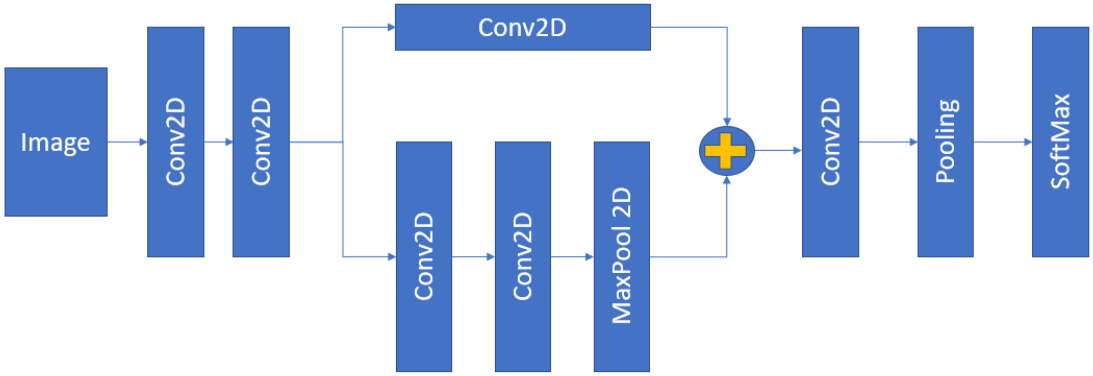


Figure 26 Architecture of Convolutional Neural Network for Real-Time Facial Expression Classification From **(Arriaga et al., n.d.)**

5.3.2 Anomaly detection of multivariate time series data

Anomaly detection is to detect anomalous data points in a dataset. There are mainly three broad categories of anomaly detection techniques exist: unsupervised anomaly detection, supervised anomaly detection, and semi-supervised anomaly detection (G. Li et al., 2017). Unsupervised anomaly detection detects anomalies in an unlabeled test dataset by identifying the instances which deviate from the majority of the dataset. Supervised anomaly detection techniques need a dataset with binary labels ("normal" and "abnormal"). For the testing data, the model can generate the likelihood of the testing data being normal or not. To achieve the real-time anomaly detection of ATCs' behavior, the researchers used unsupervised long short-term memory neural network to detect anomalies.

Long short-term memory neural network (LSTM) networks are recurrent neural network (RNN) models that are applicable for many time-series data processing such as speech

recognition (Hochreiter & Schmidhuber, 1997). Different from traditional RNNs, LSTM can perform tasks involving long term memory storage. Long term storage refers to remembering information for long periods. RNNs can learn complex temporal dynamics by mapping input sequences to a sequence of hidden states, and finally to the outputs of the neural networks. However, there exists a vanishing and exploding gradient problem when RNNs perform long-term tasks. Different from traditional RNNs, LSTM can perform tasks that involve long term memory storage. Within LSTM models, three gates exist for controlling and updating the constant error flow through the internal states of special units called “cells”: input gate, forgetting gate, and output gate. Therefore, LSTM is capable of accurately modeling complex multivariate sequences using its memory cell (Ding et al., 2018).

5.3.2.1 LSTM-based anomaly detection

Long short term memory (LSTM) was developed by (Hochreiter & Schmidhuber, 1997) to capture the temporal dependencies between serial data. In recent years, LSTM has been widely applied in different domains, such as natural language processing and activity recognition (Cai et al., 2019; Greenwood et al., n.d.; Hu et al., 2016). In this study, the author of this dissertation adapted LSTM to conduct the anomaly detection of ATCs’ behaviors. The researchers used the time series data of eye blink, head pose, and facial expressions as input. The trained LSTM can predict the values in the next few steps. By comparing the ground-truth value, the algorithm can calculate the estimation error. If the estimation error exceeds the defined threshold, the corresponding timestep will be considered as an anomaly.

5.3.2.2 Multivariate Gaussian distribution

Although the LSTM can predict the ATCs' performance in the form of feature vectors, the researchers still need to further process the data due to the lack of label of normal behaviors. In airspace monitoring, it is impractical to label the video as normal or abnormal. Relevant research in computer vision is fatigue detection. The researchers can label the video to train the model because fatigue is clear to recognize and identify by related behaviors (Ameen, 2014; Ji et al., 2004; Sigari et al., 2013). In this research, the researchers proposed to use multivariate Gaussian distribution to determine the periods which have the highest probability of being abnormal.

$$P(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^{T}\right\} \quad \text{Equation 22}$$

In $P(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}}|\Sigma|^{-1/2}\exp\ -\frac{1}{2}(x-\mu)\Sigma-1(x-\mu)^{T}$ Equation 22, $\sum$ represents the covariance matrix of the input vector, and μ is the mean vector, x is the input error vector generated from LSTM prediction.

5.4 Experiments

5.4.1 Experiment design

The researchers designed the simulation experiments and hired retired air traffic controllers to collect the data. The participants are retired ATCs with previous experience at a Terminal Radar Approach Control Facilities (TRACON). Further, the experiment has three different scenarios which have various workload. The varied workload can represent ATCs' behaviors when they confronted with different tasks. In this way, the

researchers can collect a wide range of audio data, biometric data, and how ATCs' behaviors change. To be more specific, the three scenarios include baseline scenario, high workload nominal, and high workload off-nominal. The baseline scenario has a moderate workload, around 4-5 aircraft showed up at once in the ATCs' airspace. The high workload nominal scenario has a high workload with 10-12 aircraft showed up at once. The high workload off-nominal scenario has the same air traffic density as the high workload nominal scenario. The difference is that the researchers add some incidents into the scenario, including moderate turbulence, pilot deviation without radio communication, runway switch, and aircraft minimum fuel. These incidents can test how ATCs' behaviors change and how operation errors might happen.

5.4.2 Data collection equipment

Figure 27 ATCs Workstations

Data collection at the TRACON Simulator at Poly campus, ASU

Figure 28 Data Collection at TRACON Simulator

The experiment facility has eight Air Traffic Management System stations. Either pilots or ATCs can use each station. The system can simulate a primary radar screen, terminal information, and radio communications. To collect the video data of the ATCs, the researchers used webcams in front of the ATCs and pilots. Moreover, the researchers recorded the communication audio data between the pilots and ATCs.

5.5 Results

There were three scenarios designed to explore how facial expression, head pose and eye blink changed according to traffic density: a baseline scenario with a moderate amount of traffic (the Baseline case), and two where the arrival traffic increased. The researchers used videos of the ATCs in the Baseline scenario to test the developed methodology. The future work will be testing the developed methods in the other two high-density situations.

The tested video recorded the controller for 25 minutes at 30 frames per second. Table 7 shows only part of the extracted facial expression, head pose, and eye blinks because the whole video has over 45000 frames. Pose_Rx, pose_Ry, and pose_Rz represents the pose angles of the head around the x, y, z-axes. The facial expression could be blank since there are some frames the facial expression classification model failed to detect the face reliably.

Table 7 Part of the Extracted Facial Expression, Head Pose, and Eye Blinks

frame	Timestamp (s)	pose_Rx	pose_Ry	pose_Rz	eyeblink	expression
1	0	-0.359	0.48	0.105	0	Null
2	0.033	-0.391	0.379	0.114	0	Null
3	0.067	-0.462	0.317	0.126	0	Null
4	0.1	-0.511	0.281	0.135	0	Null
5	0.133	-0.544	0.255	0.139	0	Null
6	0.167	-0.554	0.237	0.142	0	Null
7	0.2	-0.563	0.232	0.145	0	Null
8	0.233	-0.566	0.229	0.145	0	Null
9	0.267	-0.563	0.23	0.147	0	Null
10	0.3	-0.562	0.229	0.149	0	Null
11	0.333	-0.561	0.226	0.151	0	Null
12	0.367	-0.554	0.222	0.152	0	Null
13	0.4	-0.543	0.213	0.152	0	Null
14	0.433	-0.51	0.197	0.14	0.08	Neutral
15	0.467	-0.473	0.197	0.12	0.41	Null
16	0.5	-0.407	0.207	0.095	0.85	Null
17	0.533	-0.363	0.21	0.075	1.28	Null
18	0.567	-0.323	0.222	0.065	1.38	Sad
19	0.6	-0.274	0.227	0.062	1.24	Fear
20	0.634	-0.234	0.225	0.062	0.93	Null
21	0.667	-0.203	0.222	0.063	0.56	Fear
22	0.7	-0.181	0.215	0.064	0.42	Fear
23	0.734	-0.157	0.207	0.064	0.3	Happy
24	0.767	-0.143	0.192	0.067	0.48	Happy
25	0.8	-0.143	0.187	0.074	0.57	Neutral
26	0.834	-0.145	0.186	0.08	0.62	Neutral
27	0.867	-0.134	0.186	0.085	0.55	Neutral
28	0.9	-0.126	0.186	0.091	0.36	Null
29	0.934	-0.119	0.188	0.097	0.33	Neutral
30	0.967	-0.116	0.184	0.103	0.32	Neutral

To conduct the anomaly detection, the researchers transformed the facial expression from categorical data to numerical data using one-hot encoding. A one-hot encoding is a representation of categorical variables as binary vectors (K. K. Yang et al., n.d.). **Error! Not a valid bookmark self-reference.** showed the example of this one-hot encoding. The first column depicts the facial expression categories, and the second to eighth columns present the existence of the facial expression using 1 or 0. In this way, for example, the facial expression 'Angry' can be described as a vector (1,0,0,0,0,0,0).

Table 8 Transformation of Categorical Facial Expression to Numerical Values for Machine Learning

Facial Expression	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	1	0	0	0	0	0	0
Disgust	0	1	0	0	0	0	0
Fear	0	0	1	0	0	0	0
Happy	0	0	0	1	0	0	0
Neutral	0	0	0	0	1	0	0
Sad	0	0	0	0	0	1	0
Surprise	0	0	0	0	0	0	1
Null	0	0	0	0	0	0	0

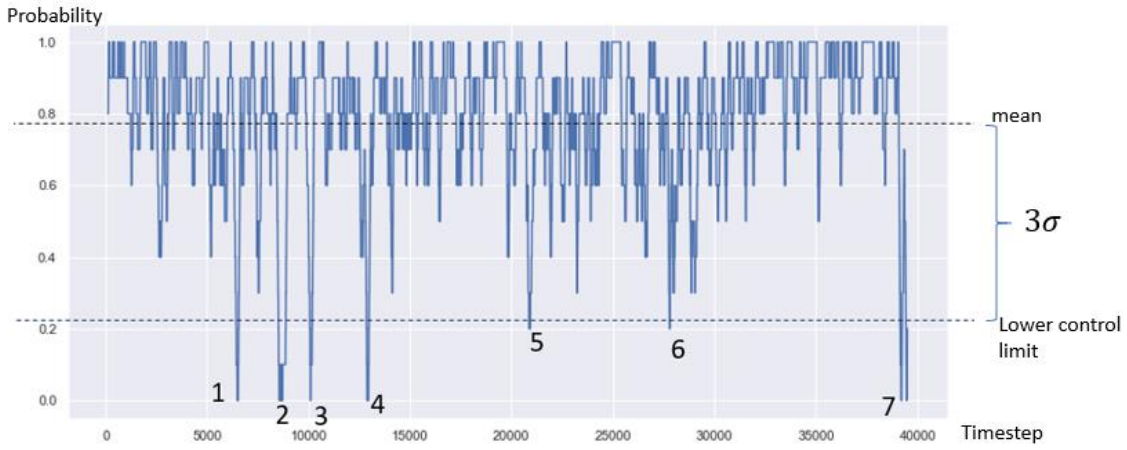


Figure 29 Probability of ATCs' Behaviors Being Normal

Table 9 Communication Analysis for Detected Anomalies (Duration: 25 Minutes; Total Words: 1703; Total Number of Messages: 138; Average Number of Messages per Minute = 5.52; Average Number of Words per Message = 12.34)

Facial Expression		**Communication**			
Anomaly	Period	Speak too fast	The message contains too much information	Missing key information	Others
1	06:56 - 07:00	4.84 < Avg.	16.00 > Avg.	Wrong call sign to Hawaiian 36: "Hawaiian thirty."	Wrong contact number:
2	08:06 - 08:18	4.76 < Avg.	23.50 >Avg.	N/A	Ambiguous clearance
3	08:57 - 09:01	10.53 > Avg.	11.25 < Avg.	Missing call sign to "Allegiant 4529)	N/A
4	10:30 - 10:35	6.00 > Avg.	17.33 > Avg.	N/A	N/A
5	14:57 - 15:00	9.38 > Avg.	11.67 < Avg.	N/A	Repeated clearance
6	18:48 - 18:49	6.45 > Avg.	13.50 > Avg.	N/A	Repeated clearance

As shown in Figure 29, the researchers used multivariate Gaussian distribution to calculate the probability of each frame being abnormal. The developed algorithm identified seven potential anomalous behaviors of ATCs at a confidence level of 99.7%. To validate the detected anomalous behaviors, the researchers analyzed the audio data to check the operations of ATCs at corresponding timeslots. As shown in Table 9, the ATCs had different communication errors in the identified anomalous periods.

Furthermore, the author of this dissertation analyzed the contextual data, which is communication data, in this case, to identify all the communication errors that happened. Figure 30 showed the transcripts to record the communication between ATCs and pilots. We identified 12 communication errors happened in the audio data, including repeated clearance, giving wrong key information, missing call sign, and ambiguous clearance. As seen in Figure 31, among the entire communication errors, 50% overlapped with the anomalies identified from video data. For the anomalies identified from video data, 83.3% of them could signify the corresponding communication errors.

	speaker	start time	transcription	end time	Duration
2	ATC (FedEx 583)	00:05.4	FedEx five eighty three, maintain four thousand, contact appraoch one two zero point niner	00:08.6	00:03.3
3	Fedex 583	00:10.7	Roger maintain four throusand, contact appraoch FedEx five eighty three	00:14.2	00:03.5
4	JetBlue 475	00:24.4	Phx approach jetblue four seventy five with you we got tango we are at one seven thousand	00:28.3	00:04.0
5	ATC (JetBlue 475)	00:29.8	jetblue four seventy five Phx approach fly heading two niner zero vector ILS? Runway two five left appraoch	00:35.5	00:05.7
6	JetBlue 475	00:36.4	heading two niner zero vector ILS? Runway two five left appraoch jetblue four seventy five	00:44.2	00:07.7
7	ATC (Southwest 1130)	01:25.4	Southwest eleven thiry desceding maintain niner thousand	01:27.3	00:01.8
8	Southwest 1130	01:29.2	Roger desceding maintain niner thousand Southwest eleven thirty	01:31.4	00:02.1

Figure 30 Part of Transcripts Used for Communication Data Analysis

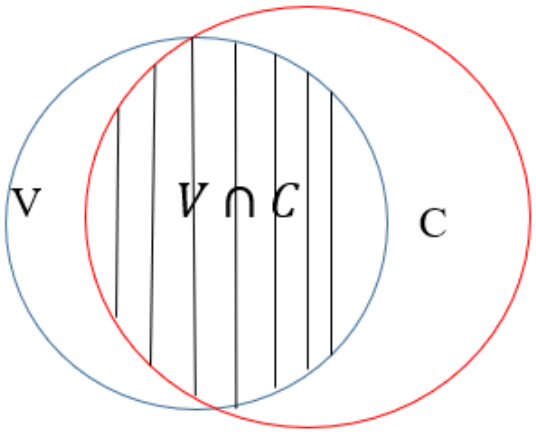


Figure 31 Anomalies Identified from Video Data and Communication Data (V Stands for Changes in Human Behaviors in Video Data; C Stands for Communication Errors).

$$\boldsymbol{V \cap C = 50\,\% C} \quad \boldsymbol{V \cap C = 83\,\% V}$$

5.6 Discussion

This research developed a real-time methodology to detect the anomalous behaviors of ATCs without a baseline of normal behaviors. Although the results showed that the developed algorithm could help monitor ATCs' performance in a non-intrusive manner. There are still some limitations, and the researchers will continue to explore in future work.

5.6.1 Implications

This section describes the potential implications of the developed human behavioral anomaly detection to engineering practices. This developed video-based anomalous behavior monitoring system could improve in-time performance monitoring of ATCs' behaviors. Different from biometric sensors, this camera-based approach could work in a non-intrusive way. For the construction domain, such research could help monitor

workers’ behaviors, such as the tower crane operators. Another finding is that the research found that head pose and eyeblink are more indicative of anomalous behaviors, and facial expression has less contribution for anomalous behaviors. For the computer science domain, this research identifies the potential usage of computer vision techniques for monitoring ATCs’ behaviors. Moreover, this research also identified that merely video data could limit the performance of the algorithm.

5.6.2 Limitations

The first limitation is that the data collection procedure is quite long, and the researchers will continue to collect more data to validate and improve the experiments. Until now, the project team spent over two years to hire retired ATCs and conduct various experiments. Due to the lack of qualified experiment participants, the researchers could not have a wide range of participants, which can consider how age, gender, physical conditions, and other related factors may influence the ATCs' performance.

The second limitation is that the researchers mainly used video to detect anomalous behaviors of ATCs. Other vital measures, including communication audio data, heart rate data, and task-related measures, are used to validate the detected anomalies. However, the researchers are exploring other data fusion techniques to include more features to identify anomalous behaviors. How including other relevant measures influence the algorithm remains unknown and challenging.

5.6.3 Future work

To address the limitations described above, the researchers will improve this work along with the following directions. The first direction will be the natural language processing

technique. One primary type of behavior of ATCs is communication with pilots. However, processing and understanding the audio data remains challenging. The communication audio data between pilots and ATCs usually involves several people and professional terms used in airspace monitoring. Separating people's voices and extract semantic information will be challenging and essential.

The second direction will be applying data fusion techniques to integrate multi-source data to improve the reliability of ATCs' performance monitoring. Currently, the researchers focused on utilizing video data to detect anomalous behaviors of ATCs. However, integrating more sensors and data will enable more reliable monitoring of ATCs' performance, including situation awareness and workload (Crutchfield, 2005; Jou et al., 2013; Truschzinski et al., 2018).

5.7 Conclusion

This paper presents a real-time methodology to identify the anomalous behaviors of ATCs using computer vision. This developed approach utilizes different facial features, including head pose, eye blink, and facial expression. Different from conventional methodology using biometric sensors to monitor the mental and physical states of ATCs, this research proved the possibility of using non-intrusive computer vision techniques for real-time monitoring of ATCs' behaviors. The results showed that the developed computer vision algorithm could signify 50% of the communication errors. As for anomalies identified from computer vision algorithms, 83% overlaps with the communication errors of ATCs.

6 CONCLUSIONS AND FUTURE RESEARCH

In this research, the author aimed at using computer vision to conduct anomaly detection for civil infrastructure. The author developed an ADCV framework that can assist in anomaly detection with or without a baseline of normal behaviors. Moreover, the author conducted a comprehensive and reliable data collection to test the developed algorithm using three case studies. The following sections described the summary of research contributions, practical implications, and recommended future research directions.

6.1 Summary of Research Contributions

6.1.1 A deep learning-based canal leakage detection approach

Maintenance planning of groundwater delivery infrastructure, such as canals, requires labor-intensive field inspection for properly allocating maintenance resources to sections of water infrastructure based on their deterioration conditions. Defective canal sections have cracks where the water delivery performance degrades. In practice, canals can be tens or even hundreds of miles long. Manual canal inspections could take weeks, while could hardly achieve comprehensive water leakage assessment. Another difficulty is that most cracks are developing under the water. Without drying up the canals, inspectors could not observe underwater conditions. They would have to assess visible parts of water facilities and environments (e.g., humidity changes and vegetation growths nearby) for prioritizing canal sections in terms of leaking risks. Even experienced inspectors need much time to complete a reliable canal condition assessment.

This research presents a deep-learning approach augmented by canal inspection knowledge to achieve automated and reliable water leak detection of canal sections from

Landsat 8 satellite images. Such integration utilizes the domain knowledge of experienced inspectors in augmenting the deep-learning methods for more reliable image pattern classification that supports rapid canal condition assessment. Compared with machine learning algorithms trained by raw satellite images manually labeled as leaking, domain-knowledge-augmented deep learning algorithms use satellite image augmented by pixel-level land surface temperature (LST), fractional vegetation coverage (FVC) and Temperature Vegetation Dryness Index (TVDI) as training samples. Specifically, LST, FVC, and TVDI for each pixel are physical parameters derived from Landsat 8 satellite images by remote sensing methods. The "leaking" or "no-leaking" labels of the training samples are from the concrete surface inspection records collected during dry-ups of the canal from 2016 to 2019. Testing results on data sets collected for canals flowing through both urban and rural areas show that the proposed approach can achieve higher recall, precision, and accuracy of leak detection compared with the conventional deep-learning method. The authors also tested how different combinations of environmental features influence the performance of the algorithm. The results showed that two feature combinations: (LST, FVC) and (LST, FVC, SH) achieve the most robust performance in diverse geospatial environments.

6.1.2 A multiple object tracking method augmented by integrating contextual information for detecting anomalous workflow

Multiple object tracking using videos has gained interest in the construction industry to support real-time monitoring and control of construction sites for productivity and safety. State-of-the-art multiple object tracking algorithms could encounter various challenges in

the field, such as inaccurate detection and identity switch because of occlusions. These field challenges cause various uncertainties in the information derived from tracking results (e.g., average waiting time of workers in a workspace). Studying the failures of multiple object tracking algorithms in various scenarios is essential for assessing the decision risks related to the use of these algorithms for deriving information used by field engineers.

The author of this dissertation has developed a multiple object tracking algorithm that strengthens state-of-the-art methods found by the authors in the literature. Further, the author characterized the performance of the algorithm using video data sets collected during a nuclear power plant (NPP) outage for monitoring indoor check-in processes of multiple workers before entering a workspace. The author proposed to use contextual integration to address the difficulties posed by the crowded and dynamic construction sites. The results indicate that the video data collected on job sites involve complex interactions among human, equipment, and environmental objects, causing various challenges to the multiple object tracking algorithm. Specifically, the authors used videos collected from an indoor space of an NPP to quantify how reliable the developed algorithm could predict the average waiting times of workers in queues at different waiting areas of the check-in space. The author categorized scenarios where multiple object tracking fails and found the significant failures came from identity switching and false positive detection of workers in the mirror or shiny surfaces. The results showed that the integration of contextual information could improve the performance of multiple object tracking. For the construction research community, this research will form a

framework to assess the reliability of multiple object tracking algorithms in deriving information used by field engineers. For the computer science community, this research identified the scenarios where state-of-art visual tracking algorithms fail to motivate the development of new algorithms.

6.1.3 A computer vision-based approach for detecting abnormal behavior of ATC

The increasing traffic in the national airspace could overwhelm Air Traffic Controllers (ATCs) and compromise the airport and airside safety. Time series of facial expressions and head pose of ATCs could capture the temporal patterns of ATCs' mental and physical states, such as emotions, gaze focus (visual attention), fatigue, confusion, and overwhelming mental workload. Anomalous temporal patterns of ATCs' states could be predecessors or indicators of dangerous mistakes or improper collaborations between ATCs and pilots. Previous research used subjective measures (Task Load Index, TLX) and physiological measures (EMG, blood pressure monitors) to collect data to analyze ATCs' mental conditions and human performance. However, these methods can be intrusive and cannot achieve real-time monitoring for practical applications. Compared with biometric sensors such as electromyography (EMG), automatic recognition of facial expression and head pose using computer vision is less intrusive for real-time ATC behavior monitoring.

The goal of this paper is to investigate two research questions: 1) how real-time computer vision methods could help recover time series of facial expressions, head poses, and eye blinks of ATCs? 2) how time series analysis methods could help identify anomalous behaviors of ATCs for guiding targeted management of ATC teams during

busy air traffic operations? The research team conducted over ten simulation experiments in a Terminal Radar Approach Control (TRACON) simulator. In these experiments, twelve experienced (retired) ATCs acted as pseudo-ATCs, and graduate students majoring in aviation traffic management at Arizona State University (ASU) acted as pseudo pilots. The research team designed Nominal High Workload and Off-Nominal High Workload scenarios to extract the pattern of controllers' facial behaviors. The results showed that a recurrent neural network-based time series analysis method could guide the detection of anomalous behaviors of ATCs and communication patterns between ATCs and pilots based on time series extracted by computer vision algorithms from videos.

6.2 Practical Implications

This research developed a systematic framework to assist computer vision-based anomaly detection for civil infrastructure operation and maintenance. This framework can help engineers achieve more automated and reliable anomaly detection using imagery data.

The canal leakage detection algorithm can well support the canal maintenance planning. The author conducted the research project with the Salt River Project and utilized the developed algorithm to assist the canal maintenance planning. The designed algorithm can generate the potential leakage areas which can help engineers focused their efforts on targeted areas. Moreover, there are more economical and accessible satellite data resources such as Planet. The high-resolution satellite image will enable a more convenient and reliable detection of canal leaks. In the future, computer vision-based

anomaly detection will support the timely and early detection of canal leaks, which can reduce water loss and soil erosion.

The multiple object tracking-based workflow monitoring system can be useful for real-world applications. In the research project, the author implemented the designed algorithm using an application to monitor the abnormal workflow. For specific civil infrastructures that have confidential concerns, the camera tends to be an ideal choice when compared to other object tracking techniques. The developed multiple object tracking approach can track workers' trajectory in the area of interest and calculate the time workers spent in different areas. The trajectory information can be useful to monitor productivity, safety, and workflow information.

The computer vision-based approach for detecting anomalous behaviors of ATCs can support the in-time monitoring of human performance. According to previous studies and reports, human errors are the major factors that lead to aviation incidents. Previous research focused on using biometric sensors to monitor human performance. The biometric sensors are intrusive and cannot work practically. For construction operations such as crane operators, non-intrusive monitoring becomes useful. The developed computer vision approach can capture the facial expression, eye blink, and head pose to provide reliable performance monitoring. The developed method can alert operators and reduce error rates.

6.3 Recommended Future Research Directions

Timely anomaly detection is vital for civil infrastructure operation and maintenance. In this thesis, the author mainly focused on how to apply computer vision to assist in

anomaly detection of civil infrastructure. However, other emerging techniques have the potentials to address the challenges and support more reliable anomaly detection. The author explored various techniques and summarized the following directions to extend the proposed ADCV framework further. The rest of this section will describe how the relevant techniques can assist practical and feasible anomaly detection.

6.3.1 Automatic extraction and integration of contextual knowledge

As for future work, considering the massive scale of civil infrastructures across the U.S., the developed algorithms could be more practical if the algorithm can adapt to different environments and can be useful for other civil facilities. The method of integrated contextual knowledge in this dissertation relies on interviews with domain experts. The author will focus on producing an automatic approach that takes environmental conditions and other physical and technical parameters of civil infrastructure systems as "contextual knowledge" to predict anomaly in civil infrastructures (Karpatne et al., 2017).

Moreover, the author will continue to improve the water leakage detection algorithm to a more extendable anomaly detection approach. The current algorithms can classify canals as leaking and non-leaking. However, for more practical use, a more detailed classification of the severeness of the anomaly is needed (J. Li et al., 2020). More accurate anomaly detection can alarm engineers in which areas need more attention and need maintenance immediately.

6.3.2 Generalized integration methods to integrate more contextual information into computer vision algorithms

In this dissertation, the author developed an algorithm that uses a layout map to provide contextual information for improving the performance of computer vision algorithms. However, the major limitation is that this approach cannot integrate other contextual information such as project schedule and BIM model. Other contextual information, for example, the BIM model, can provide detailed 3D geometric information and physical properties of civil infrastructures (Han & Golparvar-Fard, 2017). Such contextual information can provide project schedule information, equipment information to improve the performance of computer vision algorithms. A generalized integration method can utilize various contextual information to address the limitations of the computer vision algorithms in practical application on construction sites.

6.3.3 Sensor fusion to integrate contextual data for comprehensive and reliable anomaly detection of civil infrastructure

Efficient and reliable methods of canal inspection are necessary for supporting maintenance decision-making. Emerging techniques are showing the potential to achieve economical, fast, and precise water loss mapping (Huang et al., 2010a). The authors of this dissertation will examine these techniques. The author aimed at finding inspection techniques that can achieve: 1) high data collection efficiency, 2) high data accuracy, 3) low cost of technology use, and 4) high level of detail of condition assessment (Massot-Campos and Oliver-Codina 2015). For instance, the author presents a review of imaging sensors and associated techniques in water loss mapping and underwater inspection. This

research classifies the methods into two categories, a direct approach, and an indirect approach. The direct approach refers to techniques that can measure and identify the canal cracks and underwater debris directly. The underwater camera can collect imagery data directly and identify canal cracks using image processing algorithms. The indirect approach refers to those techniques which detect water leakage by analyzing the other features in the neighborhood. The authors compare these emerging technologies regarding data collection efficiency, data accuracy, the cost of technology use, and the level of detail of condition assessment.

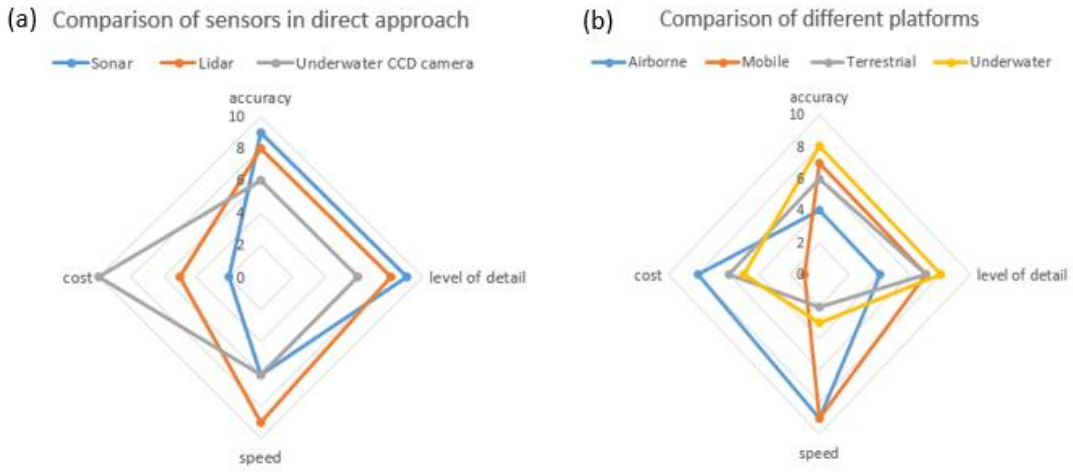


Figure 32 Comparison of Emerging Sensors and Platforms Regarding the Accuracy, Cost, Speed, and Level of Detail at A Scale of 0 – 10

Based on the comprehensive synthesis of various emerging techniques that have complementary technical characteristics, the research team found that combined use of different techniques having complementary technical capabilities would offer a better solution than the use of any single technique. The author will explore how to fuse various sensors and accomplish the project on time and within budget.

6.3.4 Cybersecurity issues of using imagery data

As mentioned above, imagery data could capture extensive spatiotemporal details (Jun Yang et al., 2015). Proper use of the imagery data becomes essential. The author of this dissertation will explore how to address cybersecurity issues of using imagery data. For example, when using video for multiple object tracking, computer vision algorithms can identify human face and characteristics, which raises privacy concerns and confidential issues (Seo et al., 2015). Masking imagery data and protecting workers' personal information will be necessary for broader applications of computer vision on construction sites.

6.3.5 Capturing and analyzing group interactions among construction workers

This thesis focused on individual performance monitoring by detecting anomalous behaviors. Construction projects generally have a large scale, and diverse activities take place there at the same time. Because most construction sites involve collaborative work, there are lots of interactions and communication between workers, workers, and machines (Jun Yang et al., 2015). Capturing and analyzing interactions among multiple groups of workers collaborating on a construction project can provide information on how workers collaborate.

The author will explore the ways to identify and monitor group behaviors so that engineers can determine which groups are more productive. The possible means of group identification can fall into four groups: shape-based, space-time, and rule-based. Shape-based methods can capture human activities with 2D or 3D skeleton pose models and recognize activities by human pose classification. Combining with the contextual

information, we can group the workers doing similar activities or relevant activities. The space-time methods identify activities based on spatiotemporal information across frames. Taking the provided literature as an example (Gudmundsson, 2017) used object tracking algorithms to extract spatial-temporal traces of player trajectories to evaluate players' performance. Rule-based approaches model activities with a set of constraints describing a set of activity patterns and recognize them by logic reasoning or pattern matching (Xiaochun Luo, Li, Cao, Dai, et al., 2018). To enable an automatic group interaction analysis, the author will implement emerging computer vision techniques such as skeleton estimation, activity recognition and instance segmentation (Lopez-Fuentes et al., 2018; Ruiz et al., 2017; Soltani et al., 2017; H. Zhang et al., 2018).

REFERENCES

ACSE. (2013). Report Card for America's Infrastructure. *American Society of Civil Engineers*, *March*, 1–74. https://doi.org/doi:10.1061/9780784478837

Adhikari, R. S., Moselhi, O., & Bagchi, A. (2014). Image-based retrieval of concrete crack properties for bridge inspection. *Automation in Construction*, *39*, 180–194. https://doi.org/10.1016/j.autcon.2013.06.011

Ameen, S. (2014). *Review of Fatigue Systems and Implementation of Face Components Segregation*. 5–8.

American Society of Civil Engineers. (2017). *2017 America's Infrastructure Report* . https://www.infrastructurereportcard.org/

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). *OpenFace: A general-purpose face recognition library with mobile applications*. http://cmusatyalab.github.io/openface/

Andriyenko, A., & Schindler, K. (2011). Multi-target tracking by continuous energy minimization. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1265–1272. https://doi.org/10.1109/CVPR.2011.5995311

Arriaga, O., Plöger, P. G., & Valdenegro, M. (n.d.). *Real-time Convolutional Neural Networks for Emotion and Gender Classification*. Retrieved November 1, 2018, from https://arxiv.org/pdf/1710.07557.pdf

Arshad, M., Gomez, R., Falconer, A., Roper, W., & Summers, M. (2014). A remote sensing technique detecting and identifying water activity sites along irrigation canals. *American Journal of Environmental Engineering and Science*, *1*(1), 19–35.

ASCE. (2013). *Report Card for America's Infrastructure*.

Ba, S. O., & Odobez, J.-M. (2009). Recognizing visual focus of attention from head pose in natural meetings. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, *39*(1), 16–33. https://doi.org/10.1109/TSMCB.2008.927274

Bailenson, J. N., Pontikakis, E. D., Mauss, I. B., Gross, J. J., Jabon, M. E., Hutcherson, C. A. C., Nass, C., & John, O. (2008). Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human Computer Studies*, *66*(5), 303–317. https://doi.org/10.1016/j.ijhcs.2007.10.011

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., & Nemani, R. (2015). *DeepSat - A Learning framework for Satellite Imagery*.

https://doi.org/10.1145/2820783.2820816

Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The CLEAR MOT metrics. *Eurasip Journal on Image and Video Processing*, *2008*. https://doi.org/10.1155/2008/246309

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. *Proceedings - International Conference on Image Processing, ICIP*, *2016-Augus*, 3464–3468. https://doi.org/10.1109/ICIP.2016.7533003

Bhandari, S., Hallowell, M. R., Van Boven, L., Gruber, J., & Welker, K. M. (2016). Emotional States and Their Impact on Hazard Identification Skills. *Construction Research Congress 2016: Old and New Construction Technologies Converge in Historic San Juan - Proceedings of the 2016 Construction Research Congress, CRC 2016*, 2831–2840. https://doi.org/10.1061/9780784479827.282

Brolly, X. L. C., Stratelos, C., & Mulligan, J. B. (2003). Model-based head pose estimation for air-traffic controllers. *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference On*, *2*, II-113–116 vol.3. https://doi.org/10.1109/ICIP.2003.1246629

Bronstein, A. M., Bronstein, M. M., Guibas, L. J., & Ovsjanikov, M. (2011). Shape Google: Geometric Words and Expressions for Invariant Shape Retrieval. *ACM Transactions on Graphics*, *30*(1), 1–20. https://doi.org/10.1145/1899404.1899405

Cai, J., Zhang, Y., & Cai, H. (2019). Two-step long short-term memory method for identifying construction activities through positional and attentional cues. *Automation in Construction*, *106*, 102886. https://doi.org/10.1016/J.AUTCON.2019.102886

Carter, S. (2015). *Controlling Water Seepage in Canals with Canal Liners*. https://www.westernliner.com/blog/control-water-seepage-with-canal-liners/

Chen, J., Tang, P., & Rakstad, T. E. (2017). Integrated Analysis of Aerial and Terrestrial Imagery Data for Efficient and Effective Water Loss Mapping of a Canal System. *Pipelines 2017: Condition Assessment, Surveying, and Geomatics - Proceedings of Sessions of the Pipelines 2017 Conference*, 136–147. https://doi.org/10.1061/9780784480885.014

Chen, Jiayu, Taylor, J. E., & Comu, S. (2017). Assessing Task Mental Workload in Construction Projects: A Novel Electroencephalography Approach. *Journal of Construction Engineering and Management*, *143*(8), 04017053. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001345

Cheng, T., Teizer, J., Migliaccio, G. C., & Gatti, U. C. (2013). Automated task-level activity analysis through fusion of real time location sensors and worker's thoracic

posture data. *Automation in Construction*, *29*, 24–39. https://doi.org/10.1016/j.autcon.2012.08.003

Crutchfield, J. M. (2005). Predicting subjective workload ratings: A comparison and synthesis of theoretical models. *ProQuest Dissertations and Theses*, *3178305*(March), 65-65 p.

Danisman, T., Bilasco, I. M., Djeraba, C., & Ihaddadene, N. (2010). Drowsy driver detection system using eye blink patterns. *2010 International Conference on Machine and Web Intelligence, ICMWI 2010 - Proceedings*, 230–233. https://doi.org/10.1109/ICMWI.2010.5648121

De la Torre, F., & Cohn, J. F. (2011). Facial expression analysis. *Handbook of Face Recognition*, 247–275. https://doi.org/10.1007/978-0-85729-997-0_19

Ding, L., Fang, W., Luo, H., Love, P. E. D., Zhong, B., & Ouyang, X. (2018). A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Automation in Construction*, *86*(October 2017), 118–124. https://doi.org/10.1016/j.autcon.2017.11.002

FAA. (2005). *Runway Incursion Trends and Initiatives at Towered Airports in the United States, FY 2001 through FY 2004*. *August*.

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., & An, W. (2018). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction*, *85*(May 2017), 1–9. https://doi.org/10.1016/j.autcon.2017.09.018

Ghanem, A. G., & AbdelRazig, Y. A. (2006). A Framework for Real-time Construction Project Progress Tracking. *Earth & Space*, *850*, 1–8.

Girardeau-Montaut, D., Roux, M., Marc, R., & Thibault, G. (2005). Change detection on points cloud data acquired with a ground laser scanner. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *36*(3), W19. https://doi.org/10.1.1.221.8313

Golabchi, A., Guo, X., Liu, M., Han, S., Lee, S., & AbouRizk, S. (2018). An integrated ergonomics framework for evaluation and design of construction operations. *Automation in Construction*, *95*, 72–85. https://doi.org/10.1016/j.autcon.2018.08.003

Golparvar-Fard, M., Asce, A. M., Feniosky, ;, Pe˜peña-Mora, P., Asce, M., & Savarese, S. (2014). *Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models*. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000205

Gong, J., & Caldas, C. H. (2010). Computer Vision-Based Video Interpretation Model for Automated Productivity Analysis of Construction Operations. *Journal of Computing in Civil Engineering*, *24*(3), 252–263. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000027

Greenwood, D., Matthews, I., & Laycock, S. (n.d.). *Joint Learning of Facial Expression and Head Pose from Speech*. https://doi.org/10.21437/Interspeech.2018-2587

Gudmundsson, J. (2017). Spatio-Temporal Analysis of Team Sports. *ACM Computing Surveys*, *50*(2), 1–34.

Gutman, G., & Ignatov, A. (1998). The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing*, *19*(8), 1533–1543. https://doi.org/10.1080/014311698215333

Hadjimitsis, D. G., Agapiou, A., Themistocleous, K., Alexakis, D. D., Toulios, G., Perdikou, S., Sarris, A., Toulios, L., & Clayton, C. (2013). Detection of Water Pipes and Leakages in Rural Water Supply Networks Using Remote Sensing Techniques. In *Remote Sensing of Environment Integrated Approaches*. http://cdn.intechopen.com/pdfs/45188/InTech-Detection_of_water_pipes_and_leakages_in_rural_water_supply_networks_using_remote_sensing_techniques.pdf

Ham, Y., Han, K. K., Lin, J. J., & Golparvar-Fard, M. (2016). Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): a review of related works. *Visualization in Engineering*, *4*(1), 1. https://doi.org/10.1186/s40327-015-0029-z

Han, K. K., & Golparvar-Fard, M. (2017). Potential of big visual data and building information modeling for construction performance analytics: An exploratory study. *Automation in Construction*, *73*, 184–198. https://doi.org/10.1016/j.autcon.2016.11.004

Hochreiter, S., & Schmidhuber, J. (1997). LONG SHORT-TERM MEMORY. *Neural Computation*.

Hu, R., Rohrbach, M., & Darrell, T. (2016). Segmentation from natural language expressions. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9905 LNCS*(d), 108–124. https://doi.org/10.1007/978-3-319-46448-0_7

Huang, Y., Associate, E., Fipps, G., & Engineer, E. A. (2008). Thermal Imaging of Canals for Remote Detection of Leaks : Evaluation in the United Irrigation District Thermal Imaging of Canals for Remote Detection of Leaks : Evaluation in the United Irrigation District 1. *Water Resources*, *November*.

Huang, Y., Fipps, G., Maas, S. J., & Fletcher, R. S. (2010a). Airborne remote sensing for detection of irrigation canal leakage. *Irrigation and Drainage*, *59*(5), 524–534. https://doi.org/10.1002/ird.511

Huang, Y., Fipps, G., Maas, S. J., & Fletcher, R. S. (2010b). AIRBORNE REMOTE SENSING FOR DETECTION OF IRRIGATION CANAL LEAKAGE. *IRRIGATION AND DRAINAGE*, *59*, 524–534. https://doi.org/10.1002/ird.511

Hunaidi, O., & Chu, W. T. (1999). Acoustical characteristics of leak signals in plastic water distribution pipes. *Applied Acoustics*, *58*(3), 235–254. https://doi.org/10.1016/S0003-682X(99)00013-4

Hunaidi, O., & Giamou, P. (1998). *GROUND-PENETRATING RADAR FOR DETECTION OF LEAKS IN BURIED PLASTIC WATER DISTRIBUTION PIPES*. http://env1.kangwon.ac.kr/leakage/2009/knowledge/papers/hardware/nrcc42068-GPRadar.pdf

Hwang, S., Asce, A. M., Jebelli, H., Asce, S. M., Choi, B., Asce, S. M., Choi, M., Asce, A. M., Lee, S., & Asce, M. (2018). *Measuring Workers ' Emotional State during Construction Tasks Using Wearable EEG*. *144*(7), 1–13. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001506.

Interior, U. S. D. of the. (2017). *Canal Operation and Maintenance : Concrete Lining and Structures*.

Inzartsev, A., & Pavi, A. (2009). AUV Application for Inspection of Underwater Communications. *Underwater Vehicles*, *December*. https://doi.org/10.5772/6704

Isaac, A., Shorrock, S. T., & Kirwan, B. (2002). Human error in European air traffic management: The HERA project. *Reliability Engineering and System Safety*, *75*(2), 257–272. https://doi.org/10.1016/S0951-8320(01)00099-0

Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technology*, *53*(4), 1052–1068. https://doi.org/10.1109/TVT.2004.830974

Jou, R. C., Kuo, C. W., & Tang, M. L. (2013). A study of job stress and turnover tendency among air traffic controllers: The mediating effects of job satisfaction. *Transportation Research Part E: Logistics and Transportation Review*, *57*, 95–104. https://doi.org/10.1016/j.tre.2013.01.009

Juba, B., Musco, C., Long, F., Sidiroglou-Douskos, S., & Rinard, M. (2015). *Long Short Term Memory Networks for Anomaly Detection in Time Series*. *April*, 22–24. https://doi.org/10.14722/ndss.2015.23268

Karpatne, A., Watkins, W., Read, J., & Kumar, V. (2017). Physics-guided Neural

Networks (PGNN): An Application in Lake Temperature Modeling. *ArXiv*. https://arxiv.org/pdf/1710.11431.pdf

Kim, D., Liu, M., Lee, S., & Kamat, V. R. (2019). Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. *Automation in Construction*, *99*, 168–182. https://doi.org/10.1016/j.autcon.2018.12.014

Kim, H., Bang, S., Jeong, H., Ham, Y., & Kim, H. (2018). Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation. *Automation in Construction*, *92*, 188–198. https://doi.org/10.1016/j.autcon.2018.04.002

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., & Fieguth, P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*, *29*(2), 196–210. https://doi.org/10.1016/j.aei.2015.01.008

Kong, J. S., & Frangopol, D. M. (2003). Life-Cycle Reliability-Based Maintenance Cost Optimization of Deteriorating Structures with Emphasis on Bridges. *Journal of Structural Engineering*, *129*(6), 818–828. https://doi.org/10.1061/(ASCE)0733-9445(2003)129:6(818)

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

Lee, Y.-J., & Park, M.-W. (2018). *3D tracking of multiple onsite workers based on stereo vision*. https://doi.org/10.1016/j.autcon.2018.11.017

Li, G., Rai, A., Lee, H., & Chattopadhyay, A. (2017). *Operational Anomaly Detection in Flight Data using a Multivariate Gaussian Mixture Model*. 1–8.

Li, J., Li, H., Umer, W., Wang, H., Xing, X., Zhao, S., & Hou, J. (2020). Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. *Automation in Construction*, *109*. https://doi.org/10.1016/j.autcon.2019.103000

Li, S., & Deng, W. (n.d.). *Deep Facial Expression Recognition: A Survey*. Retrieved September 11, 2019, from http://www.cse.oulu.fi/CMV/Downloads/Oulu-CASIA

Li, X., Wang, K., Wang, W., & Li, Y. (2010). A multiple object tracking method using Kalman filter. *2010 IEEE International Conference on Information and Automation, ICIA 2010*, *1*(1), 1862–1866. https://doi.org/10.1109/ICINFA.2010.5512258

Ligda, S. V., Seeds, M. L., Harris, M. J., Lieber, C. S., Demir, M., & Cooke, N. (2019, June 17). Monitoring Human Performance in Real-Time for NAS Safety

Prognostics. *AIAA Aviation 2019 Forum*. https://doi.org/10.2514/6.2019-3411

Liu, Y., & Goebel, K. (2018). Information Fusion for National Airspace System Prognostics. *PHM Society Conference*, *10*(1), 1–13.

Lopez-Fuentes, L., van de Weijer, J., González-Hidalgo, M., Skinnemoen, H., & Bagdanov, A. D. (2018). Review on computer vision techniques in emergency situations. *Multimedia Tools and Applications*, *77*(13), 17069–17107. https://doi.org/10.1007/s11042-017-5276-7

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Zhao, X., & Kim, T.-K. (2014). *Multiple Object Tracking: A Literature Review*. https://arxiv.org/pdf/1409.7618.pdf

Luo, X., Li, H., Cao, D., Dai, F., Seo, J., & Lee, S. (2018). Recognizing Diverse Construction Activities in Site Images via Relevance Networks of Construction-Related Objects Detected by Convolutional Neural Networks. *Journal of Computing in Civil Engineering*, *32*(3), 1–16. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000756

Luo, Xiaochun, Li, ; Heng, Cao, D., Dai, F., Asce, M., Seo, J., & Lee, S. (2018). *Recognizing Diverse Construction Activities in Site Images via Relevance Networks of Construction-Related Objects Detected by Convolutional Neural Networks*. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000756

Luo, Xiaochun, Li, H., Cao, D., Yu, Y., Yang, X., & Huang, T. (2018). Towards efficient and objective work sampling: Recognizing workers' activities in site surveillance videos with two-stream convolutional networks. *Automation in Construction*, *94*(July), 360–370. https://doi.org/10.1016/j.autcon.2018.07.011

Martin, C. A., & Gates, T. K. (2014). Uncertainty of canal seepage losses estimated using flowing water balance with acoustic Doppler devices. *Journal of Hydrology*, *517*, 746–761. https://doi.org/10.1016/j.jhydrol.2014.05.074

Milan, A., Leal-Taixe, L., Reid, I., Roth, S., & Schindler, K. (2016). MOT16: A Benchmark for Multi-Object Tracking. *IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*, 1–12. http://arxiv.org/abs/1603.00831

Mitra, V., Wang, C. J., & Banerjee, S. (2006). Lidar detection of underwater objects using a neuro-SVM-based architecture. *IEEE Transactions on Neural Networks*, *17*(3), 717–731. https://doi.org/10.1109/TNN.2006.873279

Mohd Ali, A., & Angelov, P. (2018). Anomalous behaviour detection based on heterogeneous data and data fusion. *Soft Computing*, *22*, 3187–3201. https://doi.org/10.1007/s00500-017-2989-5

Moon, W.-C., Yoo, K.-E., & Choi, Y.-C. (2011). Air Traffic Volume and Air Traffic

Control Human Errors. *Journal of Transportation Technologies*, *01*(03), 47–53. https://doi.org/10.4236/jtts.2011.13007

Mosier, K. L., Rettenmaier, P., McDearmid, M., Wilson, J., Mak, S., Raj, L., & Orasanu, J. (2013). Pilot-ATC Communication Conflicts: Implications for NextGen. *International Journal of Aviation Psychology*, *23*(3), 213–226. https://doi.org/10.1080/10508414.2013.799350

Murphy-Chutorian, E., & Trivedi, M. M. (n.d.). *Head Pose Estimation in Computer Vision: A Survey*. https://doi.org/10.1109/TPAMI.2008.106

Nealley, M. A., & Gawron, V. J. (2015). The Effect of Fatigue on Air Traffic Controllers. *International Journal of Aviation Psychology*, *25*(1), 14–47. https://doi.org/10.1080/10508414.2015.981488

Nellis, M. D. (1982). Application of Thermal Infrared Imagery to Canal Leakage Detection. In *REMOTE SENSING OF ENVIRONMENT* (Vol. 12). https://ac.els-cdn.com/0034425782900554/1-s2.0-0034425782900554-main.pdf?_tid=819cba14-af58-4342-aeef-0af4c04ae357&acdnat=1544401555_f32ba535b8da7b30664059cbc136316c

Operation, C. (2017). *Canal Operation and Maintenance : Concrete Lining and Structures*. *November*.

Ozden, A., Faghri, A., Li, M., & Tabrizi, K. (2016). Evaluation of Synthetic Aperture Radar Satellite Remote Sensing for Pavement and Infrastructure Monitoring. *Procedia Engineering*, *145*, 752–759. https://doi.org/10.1016/j.proeng.2016.04.098

Paper, C. (2016). *Leak Detection from the Buried Water Transmission Pipeline Using Landsat 8 Satellite Images (Case Study of the Kosar Water Transmission Pipeline) Leak Detection from the Buried Water Transmission Pipeline Using Landsat 8 Satellite Images (Case Study of*. *October*.

Park, H. S., Wang, Y., Nurvitadhi, E., Hoe, J. C., Sheikh, Y., & Chen, M. (2013). 3D Point Cloud Reduction Using Mixed-Integer Quadratic Programming. *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 229–236. https://doi.org/10.1109/CVPRW.2013.41

Park, M.-W., & Brilakis, I. (2016). *Continuous localization of construction workers via integration of detection and tracking*. https://doi.org/10.1016/j.autcon.2016.08.039

Park, M.-W., Elsafty, N., & Zhu, Z. (2015). Hardhat-Wearing Detection for Enhancing On-Site Safety of Construction Workers. *Journal of Construction Engineering and Management*, *141*(9), 04015024. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000974

Park, M. W., & Brilakis, I. (2012). Construction worker detection in video frames for initializing vision trackers. *Automation in Construction*, *28*, 15–25. https://doi.org/10.1016/j.autcon.2012.06.001

Paul Ekman. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression ... - Rosenberg Ekman - Google Books*. https://books.google.com/books?hl=en&lr=&id=fFGYs079-7YC&oi=fnd&pg=PR13&dq="Facial+action+coding+system+(facs),&ots=ipRiMgPuRh&sig=kQqlmH4AMeY7Z6xtHMVwvwagrA4#v=onepage&q="Facial action coding system (facs)%2C&f=false

Radopoulou, S. C., Asce, S. M., Brilakis, I., & Asce, M. (2016). *Automated Detection of Multiple Pavement Defects*. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000623

Radopoulou, S. C., & Brilakis, I. (2015). Patch detection for pavement assessment. *Automation in Construction*, *53*, 95–104. https://doi.org/10.1016/j.autcon.2015.03.010

Reddy, B., Kim, Y. H., Yun, S., Seo, C., & Jang, J. (2017). Real-Time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2017-July*, 438–445. https://doi.org/10.1109/CVPRW.2017.59

Redmon, J., & Farhadi, A. (n.d.). *YOLOv3: An Incremental Improvement*. Retrieved March 27, 2020, from https://pjreddie.com/yolo/.

Roberts, D., Bretl, T., Golparvar-Fard, M., & Student, P. D. (2017). Detecting and Classifying Cranes Using Camera-Equipped UAVs for Monitoring Crane-Related Safety Hazards. *Computing in Civil Engineering*.

Robicquet, A., Sadeghian, A., Alahi, A., & Savarese, S. (2016). Learning social etiquette: Human trajectory understanding in crowded scenes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9912 LNCS*, 549–565. https://doi.org/10.1007/978-3-319-46484-8_33

Ruiz, N., Chong, E., & Rehg, J. M. (2017). *Fine-Grained Head Pose Estimation Without Keypoints*. https://doi.org/10.1021/jp982447o

Saha, B. (2015). A Critical Study of Water Loss in Canals and its Reduction Measures. *International Journal of Engineering Research and Applications, Vol 5, Iss 3, Pp 53-56 (2015) VO - 5*, *5*(3), 53.

Salt River Project. (2015). 2015 SRP Annual Report. *Annual Report*.

Sarter, N. (2009). Error Types and Related Error Detection Mechanisms in the Aviation Domain: An Analysis of Aviation Safety Reporting System Incident Reports Nadine. *THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY*, *8414*(918552014). https://doi.org/10.1207/S15327108IJAP1002

Seo, J., Han, S., Lee, S., & Kim, H. (2015). *Computer vision techniques for construction safety and health monitoring q*. https://doi.org/10.1016/j.aei.2015.02.001

Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, *27*(6), 803–816. https://doi.org/10.1016/j.imavis.2008.08.005

Shiv Kumar, S., Wang, M., Abraham, D. M., Jahanshahi, M. R., Tom Iseley, ;, & Cheng, J. C. P. (2019). *Deep Learning-Based Automated Detection of Sewer Defects in CCTV Videos*. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000866

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Sigari, M. H., Fathy, M., & Soryani, M. (2013). A driver face monitoring system for fatigue and distraction detection. *International Journal of Vehicular Technology*, *2013*, 73–100. https://doi.org/10.1155/2013/263983

Soltani, M. M., Zhu, Z., & Hammad, A. (2017). Skeleton estimation of excavator by detecting its parts. *Automation in Construction*, *82*(May), 1–15. https://doi.org/10.1016/j.autcon.2017.06.023

Song, W., Mu, X., Ruan, G., Gao, Z., Li, L., & Yan, G. (2017). Estimating fractional vegetation cover and the vegetation index of bare soil and highly dense vegetation with a physically based method. *International Journal of Applied Earth Observation and Geoinformation*, *58*, 168–176. https://doi.org/10.1016/j.jag.2017.01.015

Soukupová, T. (2016). Real-Time Eye Blink Detection using Facial Landmarks. *Rimske Toplice*.

Spencer, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering*, *5*(2), 199–222. https://doi.org/10.1016/J.ENG.2018.11.030

Tajeen, H., & Zhu, Z. (2014). Image dataset development for measuring construction equipment recognition performance. *Automation in Construction*, *48*, 1–10. https://doi.org/10.1016/j.autcon.2014.07.006

Taneja, S., Akinci, B., Garrett, J. H., Soibelman, L., Ergen, E., Pradhan, A., Tang, P., Berges, M., Atasoy, G., Liu, X., Shahandashti, S. M., & Anil, E. B. (2011). Sensing

and field data capture for construction and facility operations. *Journal of Construction Engineering and Management*, *137*(10), 870–881. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000332

Truschzinski, M., Betella, A., Brunnett, G., & Verschure, P. F. M. J. (2018). Emotional and cognitive influences in air traffic controller tasks: An investigation using a virtual environment? *Applied Ergonomics*, *69*(December 2017), 1–9. https://doi.org/10.1016/j.apergo.2017.12.019

Tucciarelli, T., Criminisi, A., & Termini, D. (1999). Leak analysis in pipeline systems by means of optimal valve regulation. In *Journal of Hydraulic Engineering* (Vol. 125, Issue 3). https://doi.org/10.1061/(ASCE)0733-9429(1999)125:3(277)

Werner, P., Al-Hamadi, A., Niese, R., Walter, S., Gruss, S., & Traue, H. C. (2013). Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining challenges. *British Machine Vision Conference (BMVC), September*, 111–119. https://doi.org/10.5244/C.27.119

Wu, F., Mu, H., & Feng, S. (2015). Analysis of the Risk of Air Traffic Controllers' Fatigue Based on the SHEL Model. *ASCE ICTE*, 2951–2958.

Wu, Y., & Ji, Q. (n.d.). Facial Landmark Detection: a Literature Survey. *International Journal on Computer Vision*. https://doi.org/10.1007/s11263-018-1097-z

Xiao, B., & Zhu, Z. (2018). Two-Dimensional Visual Tracking in Construction Scenarios: A Comparative Study. *Journal of Computing in Civil Engineering*, *32*(3), 04018006. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000738

Yang, J., Arif, O., Vela, P. A., Teizer, J., & Shi, Z. (2010). Tracking multiple workers on construction sites using video cameras. *Advanced Engineering Informatics*, *24*(4), 428–434. https://doi.org/10.1016/j.aei.2010.06.008

Yang, Jun, Park, M. W., Vela, P. A., & Golparvar-Fard, M. (2015). Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future. *Advanced Engineering Informatics*, *29*(2), 211–224. https://doi.org/10.1016/j.aei.2015.01.011

Yang, K. K., Wu, Z., Bedbrook, C. N., & Arnold, F. H. (n.d.). *Data and text mining Learned protein embeddings for machine learning*. https://doi.org/10.1093/bioinformatics/bty178

Younis, S. M. Z., & Iqbal, J. (2015). Estimation of soil moisture using multispectral and FTIR techniques. *Egyptian Journal of Remote Sensing and Space Science*, *18*(2), 151–161. https://doi.org/10.1016/j.ejrs.2015.10.001

Yu, X., Guo, X., Wu, Z., Yu, X., Guo, X., & Wu, Z. (2014). Land Surface Temperature

Retrieval from Landsat 8 TIRS—Comparison between Radiative Transfer Equation-Based Method, Split Window Algorithm and Single Channel Method. *Remote Sensing*, *6*(10), 9829–9852. https://doi.org/10.3390/rs6109829

Yu, Y., Yao, H., & Liu, Y. (2019). Aircraft dynamics simulation using a novel physics-based learning method. *Aerospace Science and Technology*, *87*, 254–264. https://doi.org/10.1016/j.ast.2019.02.021

Zanganeh, R., Mojaradi, B., & Jabari, E. (2016). Leak Detection from the Buried Water Transmission Pipeline Using Landsat 8 Satellite Images (Case Study of the Kosar Water Transmission Pipeline). *International Conference on Civil Engineering*.

Zaurin, R., & Catbas, F. N. (2010). Integration of computer imaging and sensor data for structural health monitoring of bridges. *Smart Materials and Structures*, *19*(1), 015019. https://doi.org/10.1088/0964-1726/19/1/015019

Zhang, C., Tang, P., Cooke, N., Buchanan, V., Yilmaz, A., St. Germain, S. W., Boring, R. L., Akca-Hobbins, S., & Gupta, A. (2017). Human-centered automation for resilient nuclear power plant outage control. *Automation in Construction*, *82*(May), 179–192. https://doi.org/10.1016/j.autcon.2017.05.001

Zhang, H., Yan, X., & Li, H. (2018). Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Automation in Construction*, *94*(May), 1–10. https://doi.org/10.1016/j.autcon.2018.05.033

Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 22–40. https://doi.org/10.1155/2016/7954154

Zhou, Y., & Tuzel, O. (2017). VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. *ArXiv*.

Zhu, Z., Ren, X., & Chen, Z. (2017). Integrated detection and tracking of workforce and equipment from construction jobsite videos. *Automation in Construction*, *81*(April), 161–171. https://doi.org/10.1016/j.autcon.2017.05.005