23 Mar 2021

Human-Robot Collaboration for Effective Bridge Inspection in the Artificial Intelligence Era

Ruwen Qin
Stony Brook University

INSPECTING AND PRESERVING INFRASTRUCTURE THROUGH ROBOTIC EXPLORATION

INSPIRE UNIVERSITY TRANSPORTATION CENTER WEBINAR

HUMAN-ROBOT COLLABORATION FOR EFFECTIVE BRIDGE INSPECTION IN THE ARTIFICIAL INTELLIGENCE ERA

MARCH 23, 2021

Outline

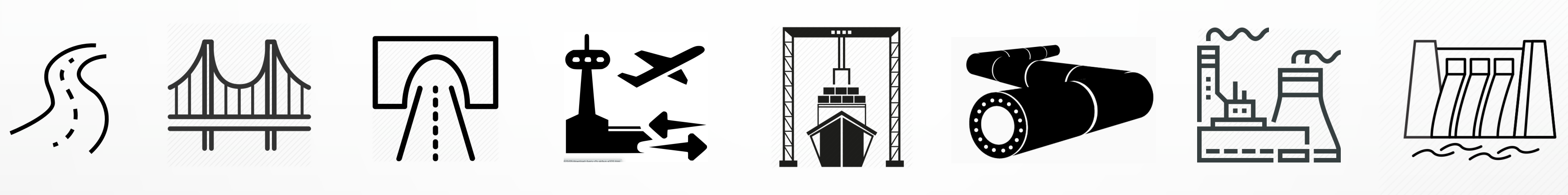
- The Problem
- Related Work
- Our Approach
- Results & Discussions
- Recommendations & Future Work

Outline

- The Problem
- Related Work
- Our Approach
- Results & Discussions
- Recommendations & Future Work

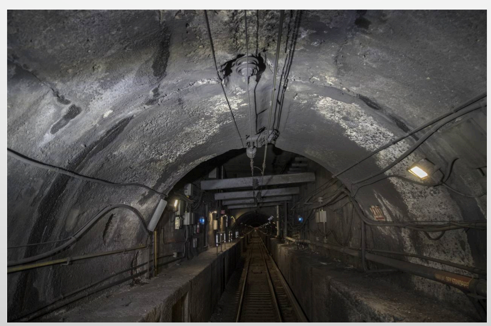
US Civil Infrastructure Systems

- Increasingly broad and more complex



- Operate in harsh environment due to meteorological and electrochemical impacts
- Subject to multiple hazards such as earthquake, flood, hurricane, tornado, and tsunami
- Deteriorate rapidly and may be approaching their designed life spans

https://www.modot.org/chariton-county-us-route-24-middle-fork-chariton-river-bridges-completed



https://www.wsj.com/articles/cuomo-digs-in-on-aging-train-tunnel-to-penn-station-telling-trump-to-fork-over-funding-1539880705

https://www.cbs17.com/news/hundreds-pack-meeting-about-woodlake-dam-in-moore-county/

Bridges

- Over 600,000 highway bridges in the National Bridge Inventory (NBI)
- 39% have exceeded their design life of 50 years
- 9% are structurally deficient and require significant repair
- National Bridge Inspection Standards require bridges to be inspected every two years to ensure that there are no cracks, rusting, or other damage
- Over hundreds of bridges need to be inspected every day

Problems of The Current Practice of Bridge Inspection

blocking the traffic

time-consuming in data collection

difficulty in accessing elevated bridge structures

dangerous field activities

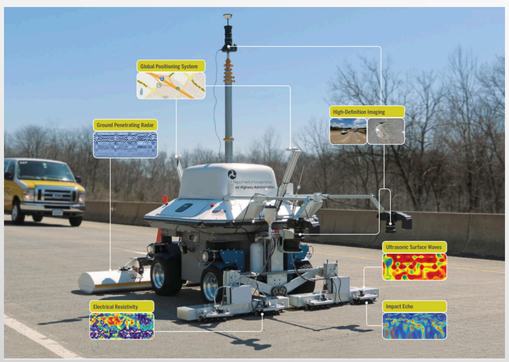
Bias of human visual inspection

Bridge Element Inspection

- Make visual inspection more objective
- Promote data-driven asset management
- AASHTO published the 2nd edition of *Manual for Bridge Element Inspection*
  - Standardized element definitions
  - Element quantity calculations
  - Condition state definitions
  - Element feasible actions
  - Inspection conventions
- Obstacles are present in executing the bridge element inspection thoroughly
  - Data collection
  - Computational tools for evaluating element conditions and the overall condition of the bridge

Robotic Technology for Bridge Inspection

- To address the cost, accessibility, safety, and reliability concerns of the current inspection practice
- To move forward to data-driven asset management
- Robotic platforms equipped with sensors



RABIT autonomous deck inspection by Rutgers Univ. - CIAT

BIRDS – Missouri S&T, Dr. Genda Chen

Climbing robot – University of Nevada, Reno, Dr. Hung La

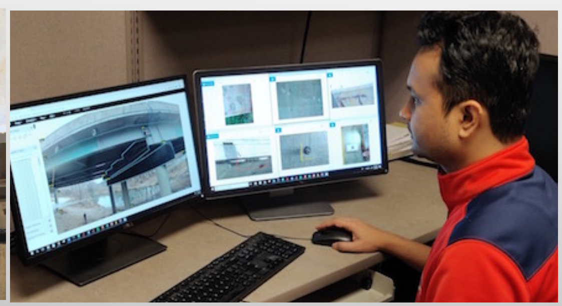
Future Work at the Human-Technology Frontier

- Current practice – Inspectors visually check the bridge conditions at the site


https://studopedia.org/1-87707.html

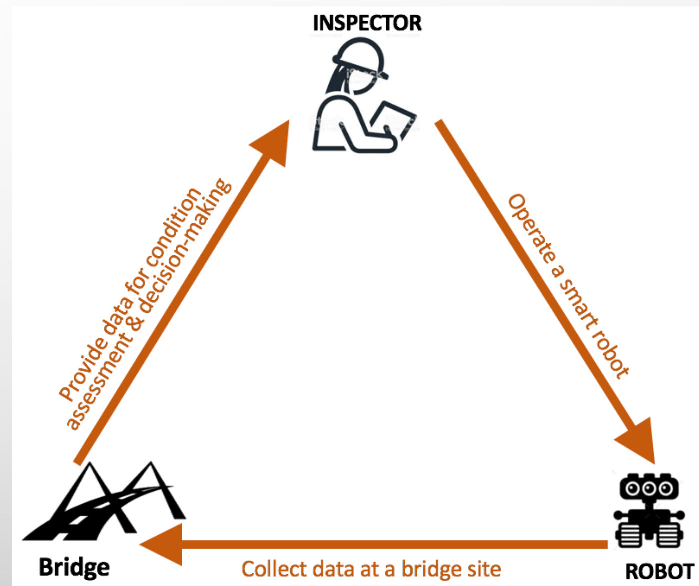
- Future work – cooperative robot-inspector survey of bridges

Dr. Genda Chen, Missouri S&T



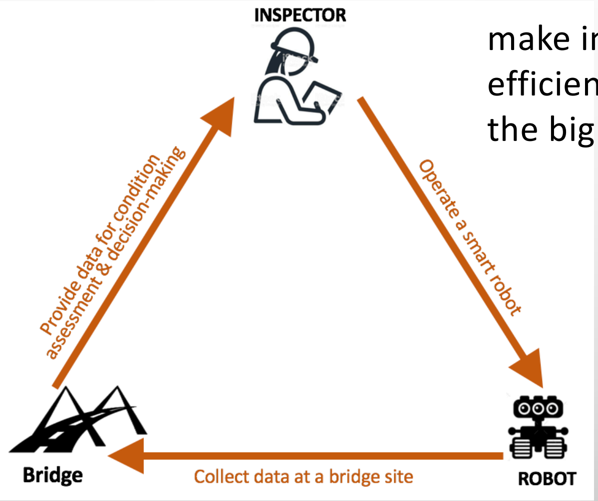
Muhammad Monjurul Karim, Stony Brook

- A paradigm shift of bridge inspection due to the advancement of new technologies



New Challenges Facing Inspector-Robot Collaboration

- Capability of the inspector-robot system is not linear additives of constituents' capabilities.
- Requires solving a series of interdependent problems.
- Focuses of Qin's group: enable inspector-robot collaboration in bridge inspection by developing computational intelligence methods and training systems



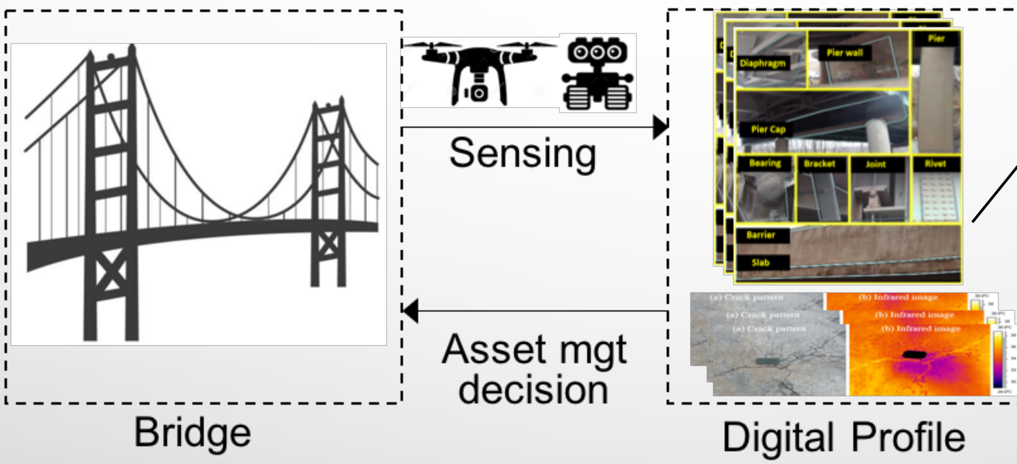
make inspectors highly confident and efficient in operating robots and analyzing the big complex image/video data.

create an accurate digital profile of every bridge, that can be updated over time, to support the data-driven asset management

maximize the utilization and efficiency of robots in assisting inspectors to collect high-quality data given the operational constraints

Assistive Intelligence for Inspection Video Data Analysis

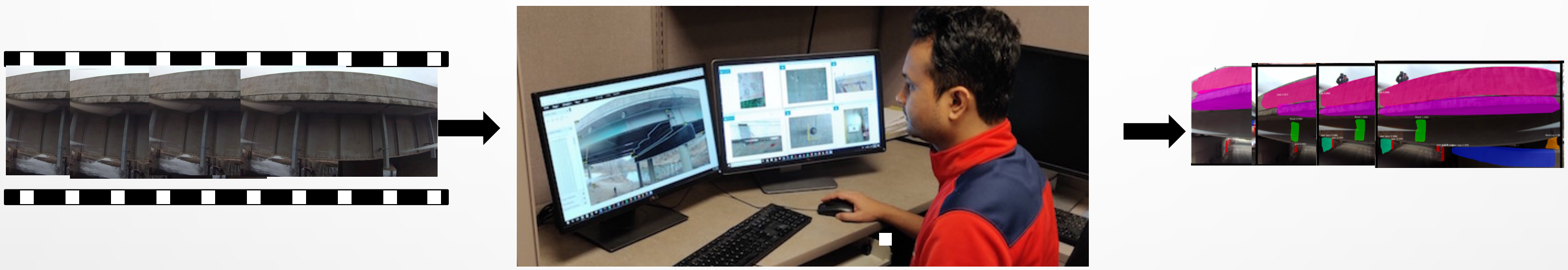
- How to assist inspectors in analyzing inspection video data efficiently and effectively



Inspection video data analysis for creating the digital profile of bridges

Inspection Video Data Analysis

1. Detect and segment bridge elements from inspection video data and sort them out by classes



1) Barrier
2) Slab
3) Bearing
4) Pier
5) Pier cap
6) Pier wall
7) Rivet
8) Truss
9) Bracket
10) Joint

2. Detect and assess defects on bridge elements to determine the condition of elements

excellent

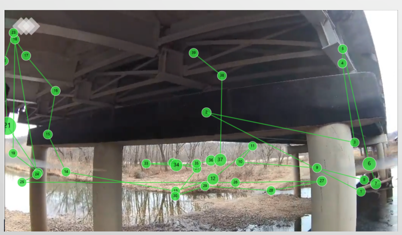
good

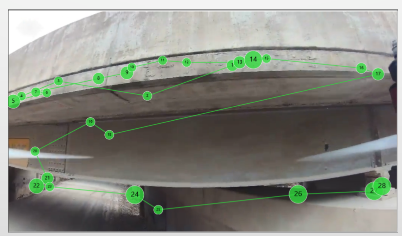
fair

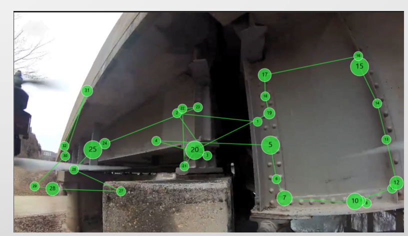
poor

AI Assisted Inspection Video Data Analysis

- Inspection video data analysis is challenging
    - Volume: a standard RGB camera collects 108K images per hour
    - Velocity: robotic inspection platforms can collect data in fast speed
    - Variety: elements of 600k bridges exhibit large variations in videos and mix with cluttered background
    - Veracity: data quality is affected by viewpoint changes, occlusion, camera vibration and distortion, scale variation, limitation of natural illumination, etc.
- Letting inspectors watch inspection videos for hours and days to locate the desired regions of interest is low in efficiency and easily triggers the development of fatigue.







- Develop computational intelligence models to assist inspectors in processing the inspection video data.

Outline

- The Problem
- Related Work
- Our Approach
- Results & Discussions
- Recommendations & Future Work

Related Work

- Bridge element segmentation from videos
  - Multiscale Siamese neural network initially trained by one-shot learning and fine-tuned iteratively with human-in-the-loop for segmenting a single class of elements (Zhao et al., 2019)
  - Convolutional neural networks for locating and extracting regions of interest (ROI) + image classification for filtering corrupted ROIs (Yeum et. al., 2019)
  - Multiscale convolutional neural networks + the scene classification for reducing false positives (Narazaki et al., 2018, 2020)
  - Segmenting multiclass structural elements from the video data collected by aerial robotic platforms for bridge element inspection is not solved completely
- Multiclass object detection & segmentnation
  - Region-based CNN (R-CNN) (Girshick et al., 2014), Fast R-CNN (Girshick et al., 2015), Mask R-CNN (He et al., 2017) for multiclass object detection and/or segmentation from static images
  - Temporal coherence information to address the issue of inconsistent detection (Kang et al., 2017; Zhu et al., 2017a, 2017b), which are computational expensive
  - Seq NMS (Han et. al., 2016) uses post-processing and is efficient. But false positives are an issue.

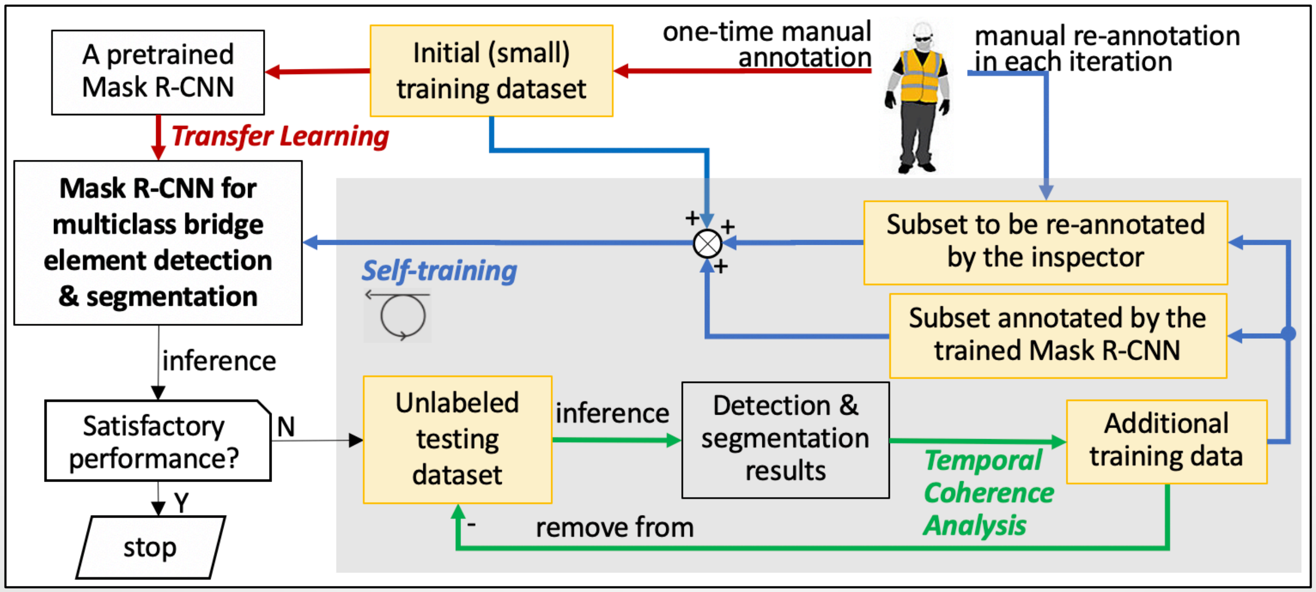
Related Work (cont.)

- Transfer learning
  - Deep learning does not solve all challenges
  - Address the issue of expensive data annotation by inspectors
  - Has been used for structural damage detection (Zhang et al., 2018; Gao and Mosalam 2018; Zhang and Chang, 2019; Gopalakrishnan et al. 2018)
- Semi-supervised learning
  - Self-training combining a portion of human annotated data and a portion of model annotated data (Triguero et al., 2015)
  - Performance is not good enough, mainly due to the quality of automatically annotated dada
- Active learning
  - Actively select training samples from a pool of unlabeled data and let a human annotator to label them for re-training the model
  - Determining the most informative samples for active learning is complex (Settles, 2012; Tian et al., 2020; Beluch et al., 2018; Sener and Savarese, 2017; Morrison et al., 2019; Gal et al. 2017; Yang et al., 2017; Siddiqui et al. 2020)
  - Simple but effective methods for recommending new data for annotation are greatly desired

Outline

- The Problem
- Related Work
- Our Approach
- Results & Discussions
- Recommendations & Future Work

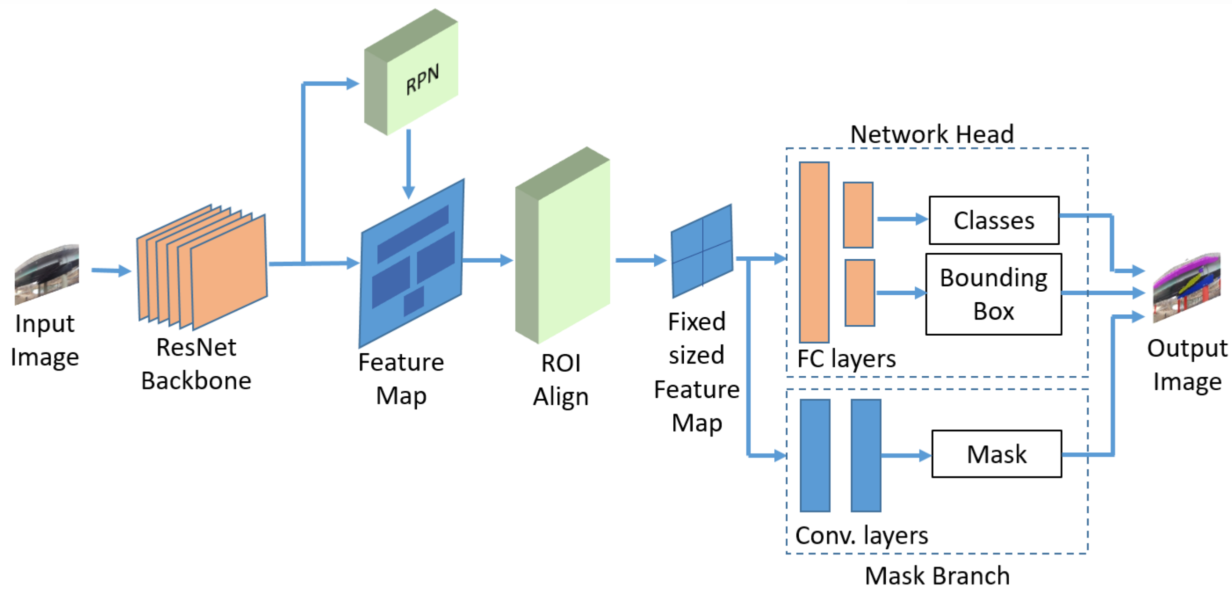
Engage Inspectors to Develop Assistive Intelligence



1. Transfer learning for the initial adaption of Mask R-CNN to the new task
2. Temporal coherence analysis that identifies the weakness that current neural network can learn from
3. Iterative self-training that engages inspectors to boost the performance

Transfer Learning

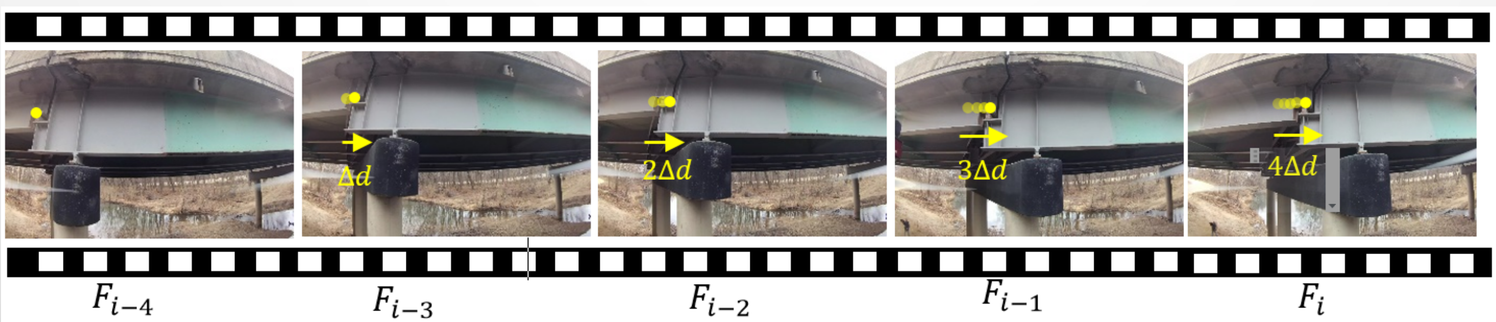
- A Mask R-CNN for segmenting multiclass bridge elements from inspection video data
- Training this deep learning network from the scratch requires a large volume of annotated data to achieve satisfied prediction accuracy



- A Mask R-CNN, pre-trained on Microsoft COCO, is transferred in for the task of multiclass bridge element detection and segmentation
- A small initial training dataset (T_0) annotated by inspectors: 40 images containing 482 labeled objects in 10 classes

Temporal Coherence Analysis

- Identify the weakness of the current network through recovering false negatives, and accordingly, recommend additional training data to retrain the network.
- An object in a frame is highly likely present in the neighboring frames within a range of spatial displacement with a similar confidence.
- Example:

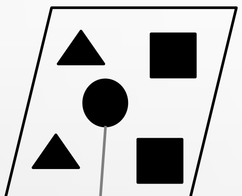


- Temporal coherence information: objects with high detection scores in preceding frames and their spatial locations
- Use temporal coherence information to recover false negatives

Temporal Coherence Analysis (cont.)

Frame $i-(q+1)$

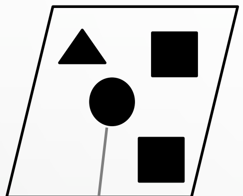
Set of detection $O_{i-(q+1)}$



A 'dot' is found within the distance $(q+1)\Delta d$ from $C_{i,j}$ with detection score $S_{i-(q+1),j'}$ and at the location $C_{i-(q+1),j'}$.

Frame $i-q$

Set of detection O_{i-q}



A 'dot' is found within the distance $q\Delta d$ from $C_{i,j}$ with detection score $S_{i-q,j''}$ and at the location $C_{i-q,j''}$.

$$S_{i,j} = \frac{1}{2}\left(S_{i-(q+1),j'} + S_{i-q,j''}\right)$$

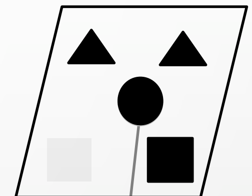
…

Search the same detection 'dot' within the surrounding area in a pair of preceding successive frames up to k frames from the nearest pair to the farthest pair (i.e., $q = 1, \dots, k-1$)

Recover this "false negative" detection by updating the detection score

Frame i

Set of detection O_i

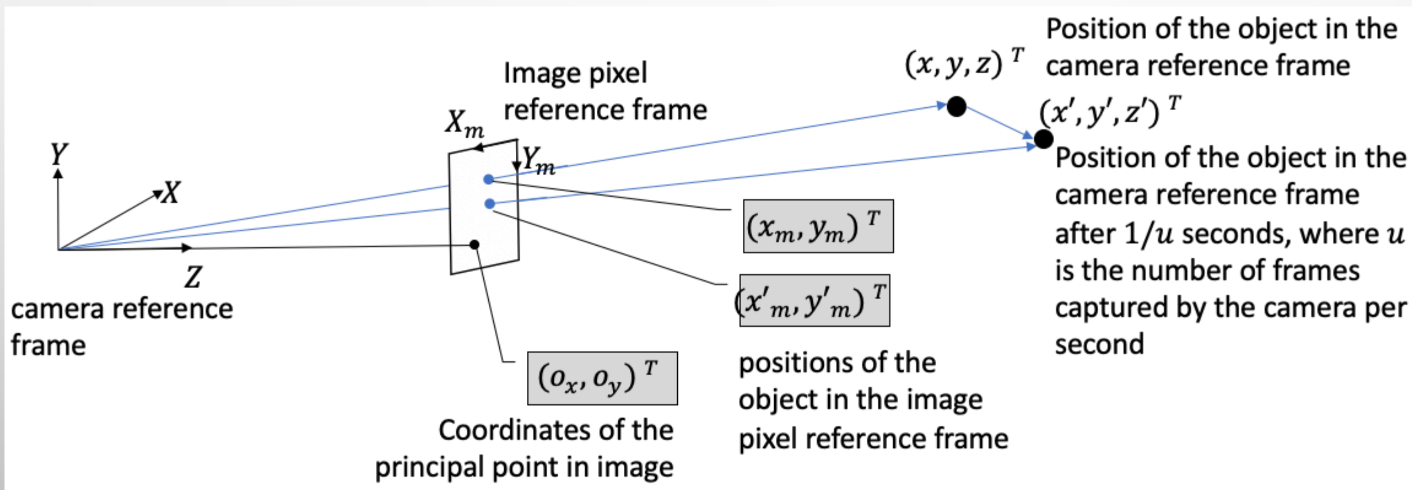


The jth object is classified as 'dot' with detection score $S_{i,j}$ and at the location $C_{i,j}$.

The detection score is lower than the upper threshold t_u of detection but greater than the lower threshold t_l, thus it may be a false negative detection.

Temporal Coherence Analysis (cont.)

- $[t_l, t_u)$ are the **threshold range** for identifying false negatives. A detection score within this range is a possible false negative.
- k is the **temporal window** that defines the range of preceding frames for searching the same objects as the false negative.
- Δd defines the **maximum spatial displacement** of objects in any two successive frames



$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} o_x \\ o_y \end{bmatrix} - \frac{f}{s_m}\begin{bmatrix} x/z \\ y/z \end{bmatrix}$$

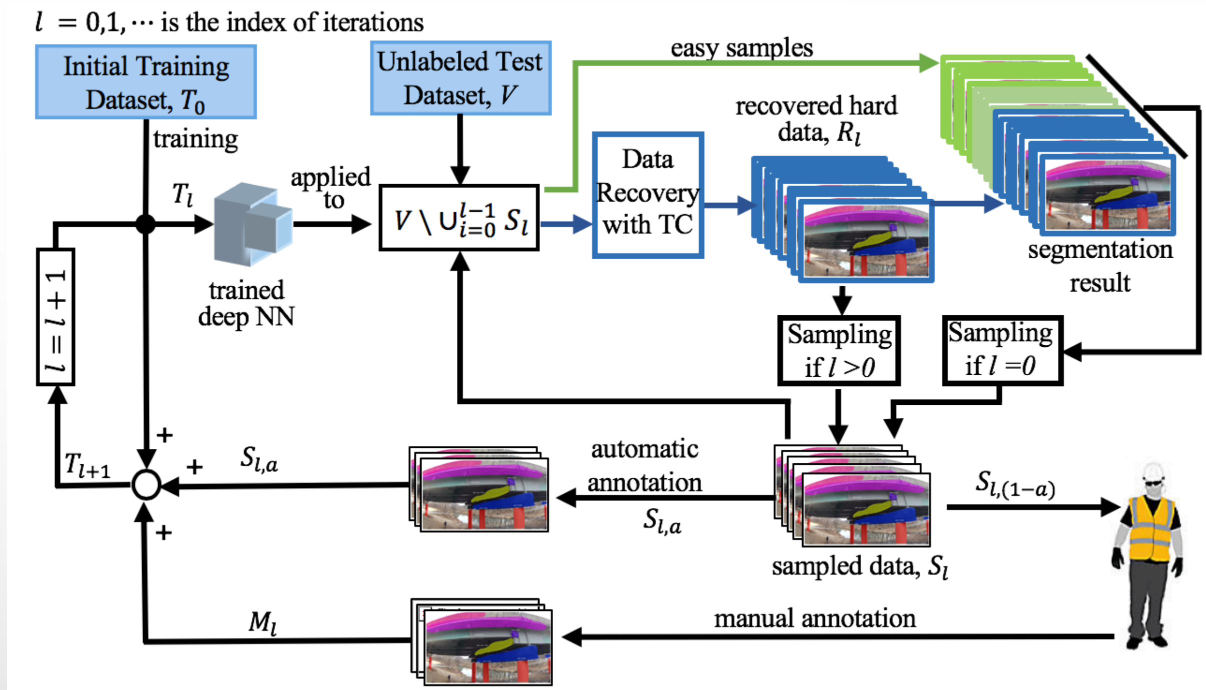
$$\begin{aligned} \Delta r &= \sqrt{(x'_m - x_m)^2 + (y'_m - y_m)^2} \\ &= \frac{f}{s_m}\sqrt{\left(\frac{x'}{z'} - \frac{x}{z}\right)^2 + \left(\frac{y'}{z'} - \frac{y}{z}\right)^2} \\ &\leq \frac{f}{s_m}\frac{\epsilon}{\min(|z|, |z'|)} = \Delta d \end{aligned}$$

f: focal length
s_m: size of pixel
$\min(z, z')$: distance from the camera to the object along the optical axis

Semi-Supervised Self-Training (S³T) with Human-in-the-Loop

- Self-training & active learning
- Sample a portion of recovered hard data that have been automatically annotated by the current network
- Let the inspector re-annotate a small portion
- Supplement the existing training dataset to retrain the network

$$T_{l+1} = T_l + M_l + S_{l,\alpha}$$



- The small dataset labeled by an experienced inspector represents the weakness of the current model and, thus, can effectively boost the performance of the network.

Semi-Supervised Self-Training (S^3T) with Human-in-the-Loop (cont.)

- Skip sampling method, SP(s), which samples a frame and then skips s frames because consecutive frames of a video are similar.

$$I_S = I_R \circ I_{SP(s)}$$

I_R: what frames are recovered hard data
$I_{SP(S)}$: what frames are chosen according to SP(s) method
I_S: what hard recovered frames are sampled

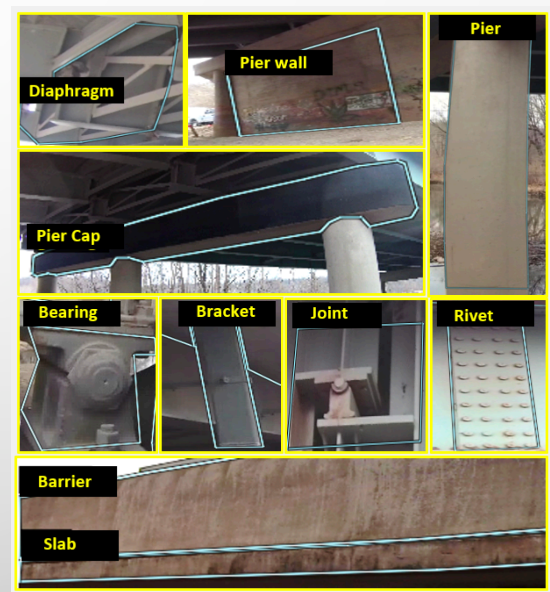
- Regulating the portion of model-annotated additional training data, α.
  - Limitation of the self-training: unsatisfied initial performance of the network causes data mislabeling
  - Letting an inspector to review a portion of the sampled data is a regulation
  - The experienced inspector guides the network to quickly learn from its weakness
  - α can be increased as the network's performance increases

Outline

- The Problem
- Related Work
- Our Approach
- Results & Discussions
- Recommendations & Future Work

Data

- Captured by the BIRDS in INSPIRE UTC: 20 mph, 30 fps, 3,840x2,160 pixel
- 10 classes of structural elements
- Dataset, D: 4,400 images from inspecting a bridge
- Initial training dataset, T_0: 40 images with 382 objects from the 10 classes with class labels
- Unlabeled dataset, V, provides additional training data during S^{3}T: 670 images with 5,916 objects from the 10 classes without class labels
- Test dataset, T_s, 212 images with 1,872 objects



The Solution Process

- Training and testing were performed using two NVIDIA Tesla V100 GPUs.
- Initial adaption of pre-trained Mask R-CNN by refining Network Head and Mask Branch
- Then, the S^3T algorithm took three iterations to complete the refining process.
- Used 4.82 hours in total (3.58 hours for data annotation and 1.24 hours for training) to develop the assistive intelligence model.

	TL	S^3T		
Index of iteration, l	0	1	2	3
Training dataset (# images), T_l	40	48	85	118
Hard recovered data (# images), R_l		113	79	50
A sampled subset of R_l (# images), S_l		37	33	
Manually annotated data (# images), M_l	8	11	7	
Automatically annotated data (# images), $S_{l,\alpha}$		26	26	
The portion of S_l for automatic annotation, α		0.7	0.8	

The Method for Evaluating Object Detection

- The bounding box IoU, IoU_B, is defined as:

$$\mathrm{IoU}_B = \frac{A_{Bp} \cap A_{Bg}}{A_{Bp} \cup A_{Bg}}$$

A_{Bp}: area of the predicted bounding box
A_{Bg}: area of the ground truth bounding box

- If the IoU_B is no less than a pre-specified IoU threshold, the prediction is a correct detection
- Measure the detection performance using Recall, Precision, and F1-score

$$\mathrm{Recall} = \frac{\#\ \mathrm{correct\ predictions}}{\#\ \mathrm{ground\ truth\ objects}} \qquad \mathrm{Precision} = \frac{\#\ \mathrm{correct\ predictions}}{\#\ \mathrm{predictions}} \qquad \mathrm{F1} = \frac{2\times\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Bridge Element Detection Results

- Compute the Recall, Precision, and F1-score using a range of IOU threshold values ranging from 0.1 to 0.9 at a step of 0.1.
- Results at the IoU threshold value 0.5 are in the table below:
  - Transfer learning achieved 80.3% Precision, 74.4% Recall, and 77.2% F1
  - S^3T: ↑ 11.5% precision, ↑ 19.4% Recall, and ↑ 15.5% F1, respectively

		TL	S^3T		
Index of iteration, l		0	1	2	3
IoU = 0.5	Precision (%)	80.3	81.7	90.7	91.8
	Recall (%)	74.4	90.3	90.1	93.6
	f1-Score (%)	77.2	85.8	90.4	92.7

- Very high threshold value is not necessary and increases false negatives

		TL	S^3T		
	Iteration, l	0	1	2	3
IoU					
0.1	Precision	87.5	86.3	94.0	93.5
	Recall	81.0	95.4	93.5	95.3
	F1-Score	84.1	90.6	93.8	94.4
0.2	Precision	87.1	85.7	93.9	93.5
	Recall	80.7	94.8	93.3	95.3
	F1-Score	83.4	90.0	93.6	94.4
0.3	Precision	86.8	85.2	93.7	93.4
	Recall	80.3	94.2	93.2	95.2
	F1-Score	83.4	89.5	93.4	94.3
0.4	Precision	84.6	84.2	93.2	93.1
	Recall	78.3	93.1	92.7	94.9
	F1-Score	81.3	88.5	93.0	94.0
0.5	**Precision**	80.3	81.7	90.7	91.8
	Recall	74.4	90.3	90.1	93.6
	F1-Score	77.2	85.8	90.4	92.7
0.6	Precision	75.9	77.4	85.7	88.5
	Recall	70.2	85.6	85.1	90.2
	F1-Score	73.0	81.3	85.4	89.3
0.7	Precision	65.4	66.6	74.6	78.1
	Recall	60.5	73.6	74.2	79.6
	F1-Score	60.5	73.6	74.2	79.6
0.8	Precision	43.7	43.8	50.1	49.0
	Recall	40.5	48.5	49.8	49.9
	F1-Score	42.1	46.0	49.9	49.5
0.9	Precision	6.3	3.0	6.7	4.6
	Recall	5.9	3.3	6.7	4.7
	F1-Score	6.1	3.1	6.7	4.7

The Method for Evaluating Instance Segmentation

- The mask IoU, IoU_M, is defined

$$\mathrm{IoU}_M = \frac{A_{Mp} \cap A_{Mg}}{A_{Mp} \cup A_{Mg}}$$

A_{Bp}: area of the predicted segmentation mask
A_{Bg}: area of the ground truth mask

- If the mask IoU is no less than a pre-specified threshold, the predicted mask is a correct prediction.
- Calculate the precision for each class

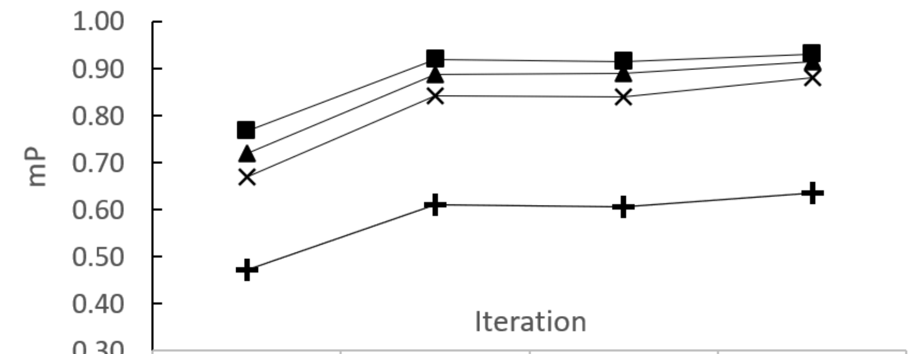
$$Pr_i = \frac{\text{\# correctly predicted segmentation mask of class } i}{\text{\# predicted segmentation masks of class } i}$$

- Average the class-level precisions to find the mean precision

$$\mathrm{mP} = \frac{\sum_{i=1}^{10} Pr_i}{10}$$

Bridge Element Segmentation Results

- Overall, mP over the iterations demonstrates an upwarding curve with reduced margin, indicating the improved segmentation performance
- Mask IoU value 0.5 is a commonly accepted minimum quality of segmentation. At this mask IoU threshold value, the developed model achieves 92% mP.
- The mP curves at threshold values 40%, 50%, and 60% are close. The curve at threshold value 70% clearly drops to a lower position.
- A very large threshold value tends to reduce the number of true positives and increase the number of false negatives.



	0	1	2	3
mP@40	0.77	0.92	0.92	0.93
mP@50	0.72	0.89	0.89	0.92
mP@60	0.67	0.84	0.84	0.88
mP@75	0.47	0.61	0.61	0.64

What Can Transfer Learning Help?

- Training the Mask R-CNN from the scratch with a training dataset of 144 images
  - the training process has not been finished after 13.2 hours (600 epochs)
  - poor performance (32.3% precision, 18.3% recall, and 23.4% F1)
- Transfer learning with the initial training dataset of 40 images
  - the training process took only 20 minutes
  - reasonable performance (80.3%, 74.4%, and 77.2%)

Method	Training time (hr)	Precision (%)	Recall (%)	F1 (%)
Training from scratch	13.2	32.3	18.3	23.4
Transfer learning	0.33	80.3	74.4	77.2

- Transfer learning allows for efficiently building an initial network for multiclass bridge element segmentation from existing deep learning neural networks for other tasks

Method Comparison

- Direct transfer learning
  - uses blind annotation. The model performance will increase at a rapidly increased cost of data annotation. But how many to annotate?
  - a performance of 89.7% precision, 92.3% recall, and 91.0% f1 took a large amount of tedious human annotation efforts (about 24 hours) plus 1.1 hours of training time to achieve.
- Self-training
  - After refine the initially transferred using additional annotated 8 images, no further guidance from the inspector

Method	No. manually annotated images	Annotation time (min)	Training time (min)	Inference speed (sec/frame)	Precision (%)	Recall (%)	F1-score (%)
Mask R-CNN	22	72	18	0.55	68.0	68.4	68.2
Mask R-CNN	44	143	20	0.55	82.0	79.0	80.5
Mask R-CNN	220	715	33	0.55	85,8	91.8	88.7
Mask R-CNN	440	1,430	66	0.55	89.7	92.3	91.0
Self-training	48	156	72	0.55	88.9	76.7	82.4
Our approach	**66**	**215**	**72**	**0.55**	**91.8**	**93.6**	**92.7**

Method Comparison (cont.)

- Compared to direct transfer learning, our approach
  - reduced the annotation time by 85% and took a comparable amount of time (only 6 more minutes) to train the model.
  - achieved a better performance (↑ 2.1% precision, ↑ 1.3% recall, ↑ 0.3% F1)
- Compared to self-training, our approach
  - Required additional one hour of data annotation by the inspector
  - Achieve a large margin (↑ 2.9% precision, ↑ 16.9% recall, ↑ 10.3% F1)

Method	No. manually annotated images	Annotation time (min)	Training time (min)	Inference speed (sec/frame)	Precision (%)	Recall (%)	F1-score (%)
Mask R-CNN	22	72	18	0.55	68.0	68.4	68.2
Mask R-CNN	44	143	20	0.55	82.0	79.0	80.5
Mask R-CNN	220	715	33	0.55	85,8	91.8	88.7
Mask R-CNN	440	1,430	66	0.55	89.7	92.3	91.0
Self-training	48	156	72	0.55	88.9	76.7	82.4
Our approach	**66**	**215**	**72**	**0.55**	**91.8**	**93.6**	**92.7**

Impact of Human-in-the-Loop to Self-training

- The S^3T method that keeps human-in-the-loop is a combination of self-training and active learning.
- The portion of additional training data re-annotated by experienced inspectors in each iteration may impact the model development efficiency and the performance of the resulting final model.
- Experimental studies
  - Experiments 1~3 are self-training with human-in-the-loop
  - Experiment 4 is conventional self-training
- Including the inspector's guidance in self-training is critical
- Following the suggested regulation assures the convergence to satisfied performance in a few iterations

	iterations	0	1	2	3
1	precision (%)	80.3	81.7	90.1	**91.8**
	recall (%)	74.4	90.3	90.7	**93.6**
	f1-score (%)	77.2	85.8	90.4	**92.7**
	$S_{l,\alpha}$ (frame)	0	26	26	
	M_l (frame)	8	11	7	
	α_l (%)	0	70	80	
2	precision (%)	80.3	81.7	91.8	**94.5**
	recall (%)	74.4	90.3	84.7	**91.5**
	f1-score (%)	77.2	85.8	88.1	**93.0**
	$S_{l,\alpha}$ (frame)	0	22	26	
	M_l (frame)	8	15	11	
	α_l (%)	0	60	70	
3	precision (%)	80.3	81.7	90.1	91.7
	recall (%)	74.4	90.3	90.7	92.1
	f1-score (%)	77.2	85.8	90.4	91.9
	$S_{l,\alpha}$ (frame)	0	26	28	
	M_l (frame)	8	11	3	
	α_l (%)	0	70	90	
4	precision (%)	80.3	81.7	88.3	88.9
	recall (%)	74.4	90.3	75.7	76.7
	f1-score (%)	77.2	85.8	81.5	82.4
	$S_{l,\alpha}$ (frame)	0	37	33	
	M_l (frame)	8	0	0	
	α_l (%)	0	100	100	

Job Efficiency Achieved

- A small-scale example of detecting and segmenting bridge elements from 20 images

Assistive Intelligence	**Work time (min)**	**Accuracy (%)**
without	65	100
with	0.27	93.7
change	↓99.5%	↓6.3%

- A real-world task would analyze hundreds of thousands of images.
- As the job size increases, the gap of accuracy is diminishing and can reverse due to human factors related issues (e.g., loss of attention, fatigue, etc.)
- The time saving is tremendous, in proportion to the size of the real-world task
- It is not a problem to let inspectors analyze a small portion of images for guiding the algorithm development, but not realistic to let inspectors to manually analyze all.

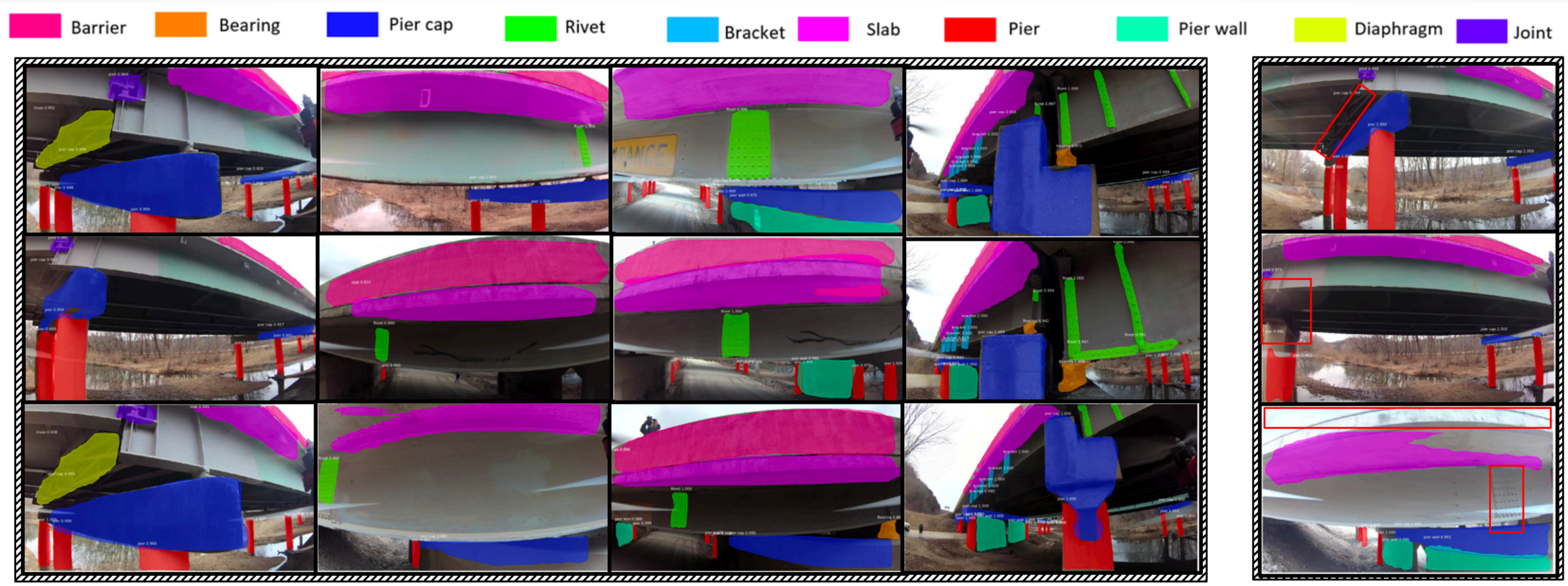
Generalization Capability of the Proposed Method

- What if we use the neural network that is developed for one bridge to inspect other bridges?
- The bridge that has had a network developed for is named bridge C.
- Segment the same ten classes of bridge elements for another two bridges, named bridges A and B.
  - Bridge A: 76.8% Precision, 73.0% Recall, and 74.8% F1-score
  - Bridge B: 61.2% Precision, 60.0% Recall, and 60.6% F1-score
- Comparable to the performance of the initial network for bridge C.
- The network developed in this study has a certain degree of generalization, and it is good enough to serve as the initial network for other bridges.
- Implementing the S^3T method developed in this study will adapt the neural network to other bridges.

Representative Examples

(a) Successful detection and segmentations

(b) False negative detections



(a) (b)

Outline

- The Problem
- Related Work
- Our Approach
- Results & Discussions
- Recommendations & Future Work

Recommendations

1. No deep learning models are directly applicable to any tasks and, therefore, it is important to adapt models to different new tasks to achieve satisfied performance. Transfer learning, semi-supervised learning, and active learning are examples of useful methods that let the algorithms learn features of a new task from a small amount of high-quality data of the task, thus adapting to that task.
2. Keeping human-in-the-loop is an important method to leverage human intelligence into the artificial intelligence algorithms. This should be achieved through a collaborative approach. That is, algorithms provide humans with its performance so that humans can figure out the weakness and provide inputs (e.g., additional training dataset annotated by humans) to the algorithms for improvement.
3. Model adaptability and the collaboration between artificial intelligence and human experts were integrated together as a solution for developing assistive intelligence that takes care of time-consuming, boring tasks and let humans focus on knowledge-intensive tasks. This will be a new style of work for future bridge professionals.

Future Work

- Adapting the assistive intelligence model to bridges with additional structural elements is the next step to extend this study
- Improving the testing speed to have the real time inference capability
- Contextual information and the spatial correlation among objects could be utilized to further improve the segmentation accuracy
- Evaluating the change in cognitive load and other psychological states of inspectors assisted by the assistive intelligence in bridge inspection.

Acknowledgement

- Financial support for this INSPIRE UTC project was provided by the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (USDOT/OST-R) under Grant No. 69A3551747126 through INSPIRE University Transportation Center (http://inspire-utc.mst.edu) at Missouri University of Science and Technology. The views, opinions, findings and conclusions reflected in this publication are solely those of the authors and do not represent the official policy or position of the USDOT/OST-R, or any State or other entity.
- Constructive feedback from peer reviewers
- Our team: Muhammad Monjurul Karim, Dr. Genda Chen, Dr. Zhaozheng Yin

References

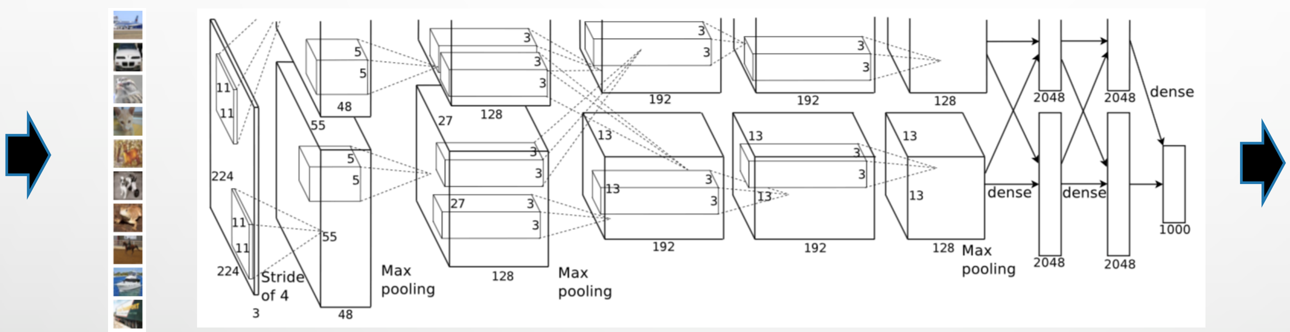
1. Zhao T, Yin Z, Qin R, Chen G. Image data analytics to support Engineers' decision making. Proceedings of the 9th International Conference on Structural Health Monitoring of Intelligent Infrastructure (2019, St. Louis, MO), Aug 2019.
2. Narazaki Y, Hoskere V, Hoang TA et al. Automated vision-based bridge component extraction using multiscale convolutional neural networks, 2018. ArXiv preprint arXiv:1805.06042.
3. Narazaki Y, Hoskere V, Hoang TA et al. Vision-based automated bridge component recognition with high-level scene consistency. Computer-Aided Civil and Infrastructure Engineering 2020; 35(5): 465–482.
4. Yeum CM, Choi J and Dyke SJ. Automated region of interest localization and classification for vision-based visual assessment of civil infrastructure. Structural Health Monitoring 2019; 18(3): 675–689.
5. Girshick R, Donahue J, Darrell T et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE 2014 Conference on Computer Vision and Pattern recognition. pp. 580–587.
6. Girshick R. Fast R-CNN. In Proceedings of the IEEE 2015 International Conference on Computer Vision. pp. 1440–1448.
7. He K, Gkioxari G, Doll´ar P et al. Mask R-CNN. In 2017 Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.
8. Kang K, Li H, Yan J et al. T-CNN: Tubelets with convolutional neural networks for object detection from videos. IEEE Transactions on Circuits and Systems for Video Technology 2018; 28(10): 2896–2907.
9. Zhu X, Xiong Y, Dai J et al. Deep feature flow for video recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2349–2358.

10. Zhu X, Wang Y, Dai J et al. Flow-guided feature aggregation for video object detection. In Proceedings of the IEEE International Conference on Computer Vision. pp. 408–417.
11. Han W, Khorrami P, Paine TL et al. Seq-NMS for video object detection. arXiv preprint arXiv:160208465 2016.
12. Zhang C, Chang C and Jamshidi M. Bridge damage detection using single-stage detector and field inspection images. arXiv preprint arXiv:181210590 2018.
13. Gao Y and Mosalam KM. Deep transfer learning for image based structural damage recognition. Computer-Aided Civil and Infrastructure Engineering 2018; 33(9): 748–768.
14. Zhang C and Chang CC. Surface damage detection for concrete bridges using single-stage convolutional neural networks. In Health Monitoring of Structural and Biological Systems XIII, volume 10972. International Society for Optics and Photonics, p. 109722E.
15. Gopalakrishnan K, Gholami H, Vidyadharan A et al. Crack damage detection in unmanned aerial vehicle images of civil infrastructure using pre-trained deep learning model. Int J Traffic Transp Eng 2018; 8: 1–14.
16. Triguero I, Garc´ıa S and Herrera F. Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. Knowledge and Information systems 2015; 42(2): 245–284.
17. Settles B. Active learning. Synthesis lectures on artificial intelligence and machine learning 2012; 6(1): 1–114.
18. Tian Y, Cheng G, Gelernter J et al. Joint temporal context exploitation and active learning for video segmentation. Pattern Recognition 2020; 100: 107158.

19. Beluch WH, Genewein T, N¨urnberger A et al. The power of ensembles for active learning in image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9368–9377.
20. Sener O and Savarese S. Active learning for convolutional neural networks: A core-set approach. In International Conference on Learning Representations. URL https://openreview.net/forum?id=H1aIuk-RW.
21. Morrison D, Milan A and Antonakos E. Uncertainty-aware instance segmentation using dropout sampling. In CVPR workshop on the Robotic Vision Probabilistic Object Detection Challenge.
22. Gal Y, Islam R and Ghahramani Z. Deep bayesian active learning with image data. In International Conference on Machine Learning. PMLR, pp. 1183–1192.
23. Yang L, Zhang Y, Chen J et al. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In International conference on medical image computing and computer-assisted intervention. Springer, pp. 399–407.
24. Siddiqui Y, Valentin J and Nießner M. Viewal: Active learning with viewpoint entropy for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9433–9443.

Does Deep Learning Solve All Challenges?

- Deep learning
  - Powerful computational resources + large-scale annotated data
  - Deep learning models such as convolutional Neural Networks (CNN) improve image analysis such as object detection and classification



Airplane
Auto
Bird
Cat
Deer
Dog
Frog
horse
Ship
Truck

AlexNet in NIPS2012

- Heavily rely on intensive, tedious efforts of human to annotate training data
- Large-scale annotated datasets do not cover bridge elements or bridge conditions
  - ImageNet: 12.4 millions of labeled training samples
  - Microsoft COCO: 10 millions of labeled training samples
- High-quality labeled training data for domain-specific applications is <u>difficult to acquire</u>

Thank you!

Q&A

ruwen.qin@stonybrook.edu

INSPECTING AND PRESERVING
INFRASTRUCTURE THROUGH ROBOTIC
EXPLORATION

INSPIRE UNIVERSITY TRANSPORTATION CENTER WEBINAR

HUMAN-ROBOT COLLABORATION FOR EFFECTIVE BRIDGE INSPECTION IN THE ARTIFICIAL INTELLIGENCE ERA

DR. RUWEN QIN

MARCH 23, 2021