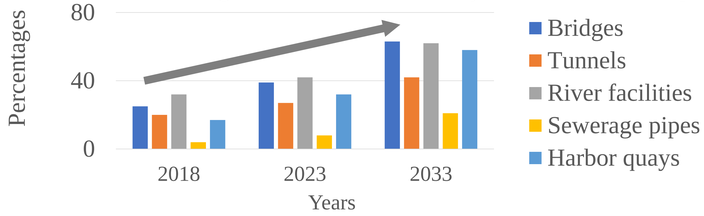
CORRELATION-AWARE ATTENTION BRANCH NETWORK USING MULTI-MODAL DATA FOR DETERIORATION LEVEL ESTIMATION OF INFRASTRUCTURES

Number: 3037

Naoki Ogawa, Keisuke Maeda , Takahiro Ogawa and Miki Haseyama
naoki@lmd.ist.hokudai.ac.jp Graduate School of Information Science and Technology, Hokkaido University

Background

Percentage of infrastructure 50 years or older in Japan [2]



Due to the increasing number of aging infrastructures, engineers need to be supported for efficient inspection.

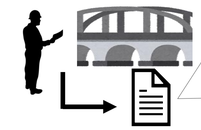
Deterioration level estimation using distress images for supporting inspections

General CNN [4]
✗ Low interpretability

Attention Branch Network (ABN) [9]
✔ Improved interpretability ✗ Low performance by using attention maps that do not highlight regions related to the estimation target [12]

Need for modification of attention maps

Engineers record text data.

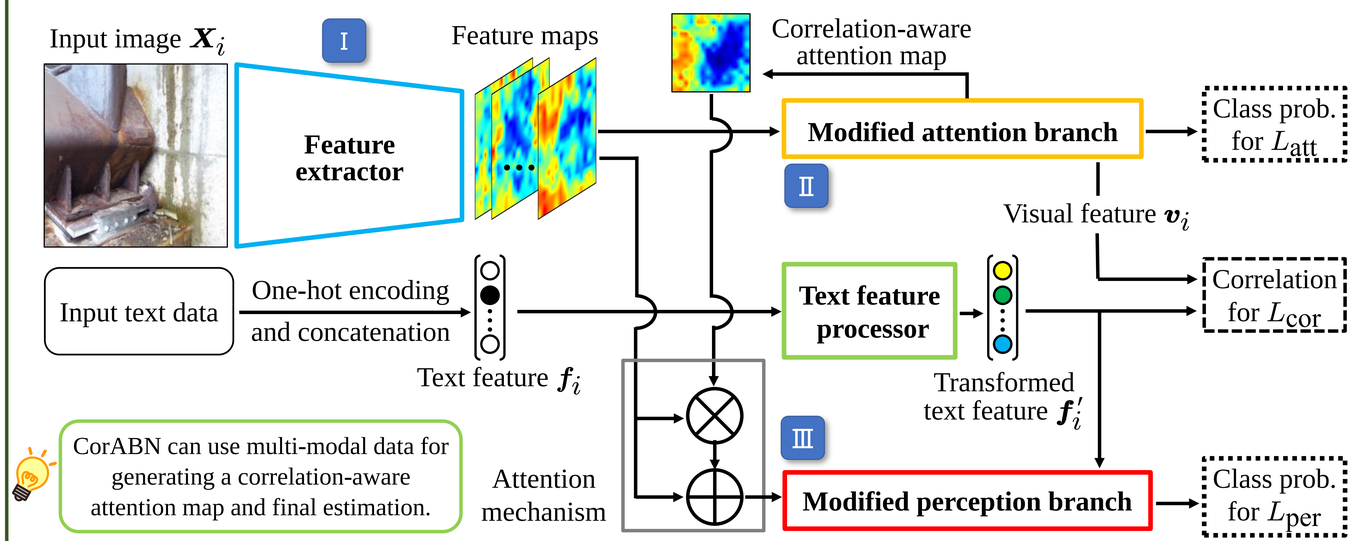


ID	Inspection items: Categories of structure	⋯	Damaged parts
1	Felloe guard	⋯	Left part, exterior part
2	RC slab	⋯	Main plate
3	Steel girder	⋯	Main girder flange
4	PC girder	⋯	Crossbeam
5	PC girder	⋯	Main girder flange

Use of text data related to distresses in distress images

Use of text data is expected to be effective in modifying attention maps.

Proposed method (PM): **Correlation-aware ABN (CorABN)**



CorABN can use multi-modal data for generating a correlation-aware attention map and final estimation.

I. Feature Extraction

Extracting features from the input image and text data

II. Generating Attention Map

Generating a correlation-aware attention map considering the correlation between visual and text features

III. Final estimation

Estimating deterioration levels using feature maps from the attention mechanism using a correlation-aware attention map

Total loss function L used for training the model

$$L = \sum_{i}^{N} \{ L_{\text{att}}(\boldsymbol{X}_i) + \alpha L_{\text{per}}(\boldsymbol{X}_i, \boldsymbol{f}'_i) + \beta L_{\text{cor}}(\boldsymbol{X}_i, \boldsymbol{f}'_i) \}$$

Losses based on the estimated class probabilities from the modified attention branch and the modified perception branch using cross-entropy loss

The loss based on the correlation between transformed text features and visual features

Training CorABN with L realizes generation of the correlation-aware attention map and accurate estimation of the deterioration level considering the text data.

* α and β are the hyperparameters.

Experimental setting

Dataset: "Corrosion" images of road infrastructures provided by East Nippon Expressway Company Limited.

Distress levels: Three classes "A", "B" and "C" in descending order of degree of distresses

The measure of performance evaluation: Macro average of f-measure (Macro-F)

Comparative methods:
CM1: AdaCos [17], CM2: ABN [9], CM3: ResNet50 [18], CM4: ResNet50 with text

Example of distress images (Corrosion):

level	Num. of images: Training	Validation	Testing
A	2,178	147	155
B	1,974	142	142
C	1,816	157	154

Results

Quantitative evaluation

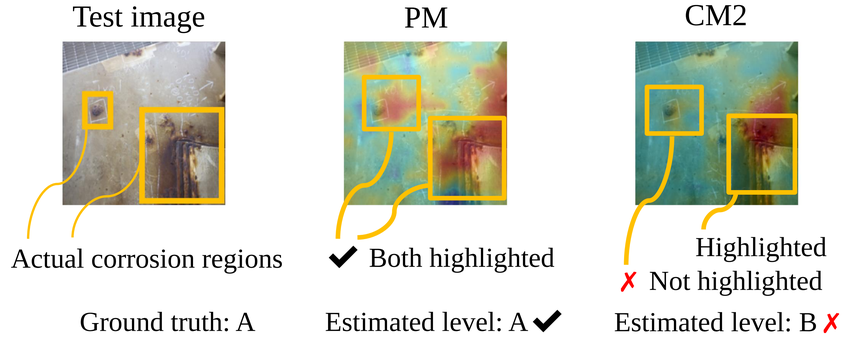
Macro-F of PM and CMs1-4

	Level A	Level B	Level C	Ave.
PM	**0.699**	**0.590**	0.698	**0.662**
CM1	0.649	0.541	**0.711**	0.634
CM2	0.644	0.511	**0.711**	0.622
CM3	0.653	0.500	0.671	0.608
CM4	0.675	0.511	0.698	0.628

PM achieves the best estimation performance.

Qualitative evaluation

In attention maps, red regions are paid attention and blue regions are not.



PM acquires the modified attention map by using text data.

Reference numbers are the same as in the paper. In this research, we used the data provided by East Nippon Expressway Company Limited. This work was partly supported by JSPS KAKENHI Grant Numbers JP17H01744 and JP20K19856.