XAI-Enhanced Semantic Segmentation Models for Visual Quality Inspection

Tobias Clement[1], Hung Truong Thanh Nguyen[1,3], Mohamed Abdelaal[2], and Hung Cao[3]

[1]Friedrich-Alexander-University Erlangen-Nürnberg, Germany
[2]Software AG, Uhlandstraße 12, 64297 Darmstadt, Germany
[3]Analytics Everywhere Lab, University of New Brunswick, Canada

Table of Content

Research Question

Current state of Explainable AI (XAI)

BEFORE

Training data → ML process → Learned Function → Decision Cat: 0.76 → End-user

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

- Explainability of XAI systems is critical for people to effectively use, interact, and achieve best outcomes with them.
- **TODAY:** XAI are typically used by **developers** for debugging and improving their systems.

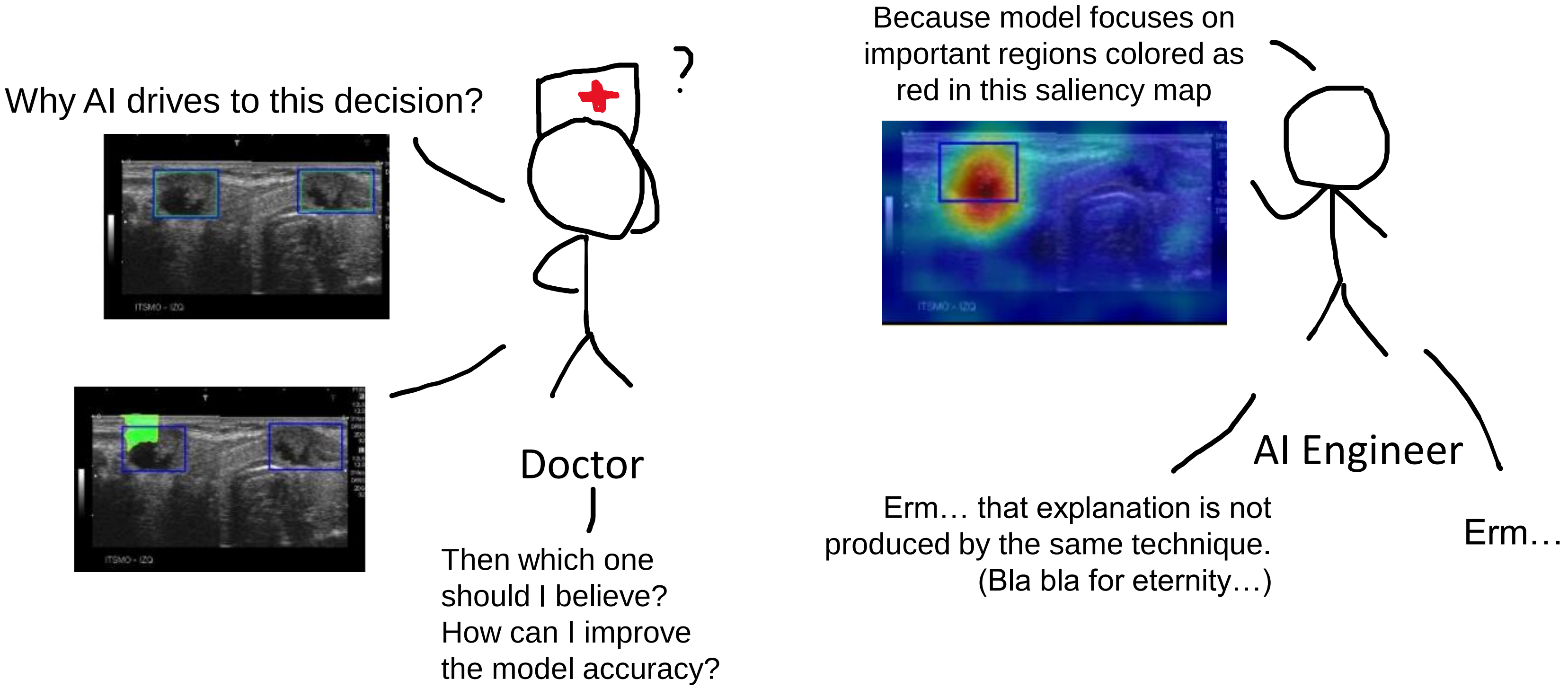
TODAY

Training data → New ML process → (Explainable) Learned Func. | Explainable decision ⇄ End-user

- I understand why
- I understand why not
- I know when you succeed
- I know when you fail
- I know when to trust you
- I know why you erred

XAI to end-users

"But little is understood about whether existing XAI explanations are understandable and useful to **end-users** of these systems who often have little to no background in AI."



Motivation

“How to deliver a XAI toolbox for semantic segmentation task with high plausible and faithful explanations to end-users and elevate the model performance with XAI?”

Use case – Visual Quality Inspection (VQI)

In our VQI, the updated field images, along with the estimates of their health, are continuously updated in the asset management information system.

This enables asset managers to plan for maintenance and replacement schedules, and gain insights into the conditions at the field level.

However, VQI may encounter several research challenges: model calibration (Rožanec et al. 2022), out-of-distribution generalization (Yang et al. 2021), and adversarial examples (Elsayed et al. 2018).

Hence, there is a demand for an enhanced VQI framework where the model can be validated, debugged, and performance-enhanced.

This necessitates the integration of XAI to provide transparency and interpretability to the decision-making processes of the models.

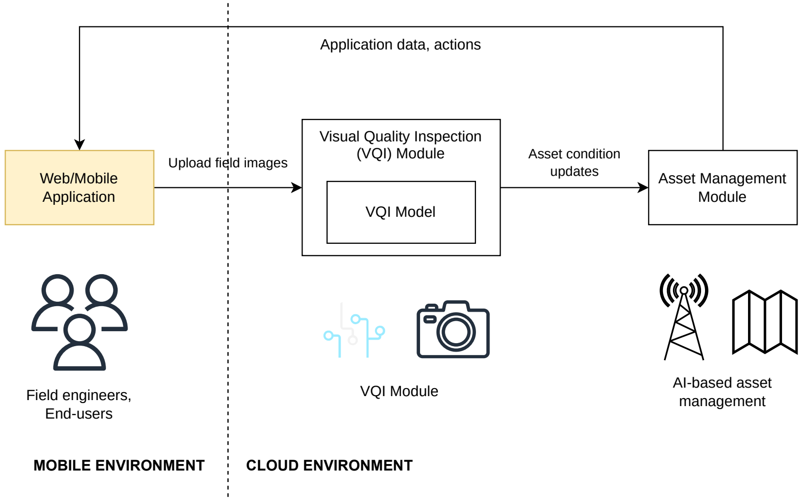


Fig 1. Conventional VQI system

Contributions

This work contributes:

1) **An Enhanced VQI Framework** by integrating XAI into the conventional semantic segmentation VQI systems, which contains 4 Building Blocks: 1 – Model Training, 2 – XAI Integration, 3 – XAI Evaluation and 4 – Model Enhancement with XAI Explanations.

2) **Evaluation procedure on XAI methods**: by using plausibility and faithfulness metrics, we can evaluate XAI methods and choose the most suitable method for the model enhancement procedure.

3) **Model Enhancement with XAI Explanations:** based on the explanation maps and suggestion of a domain expert, we can elevate the performance of semantic segmentation models through the annotation augmentation method.

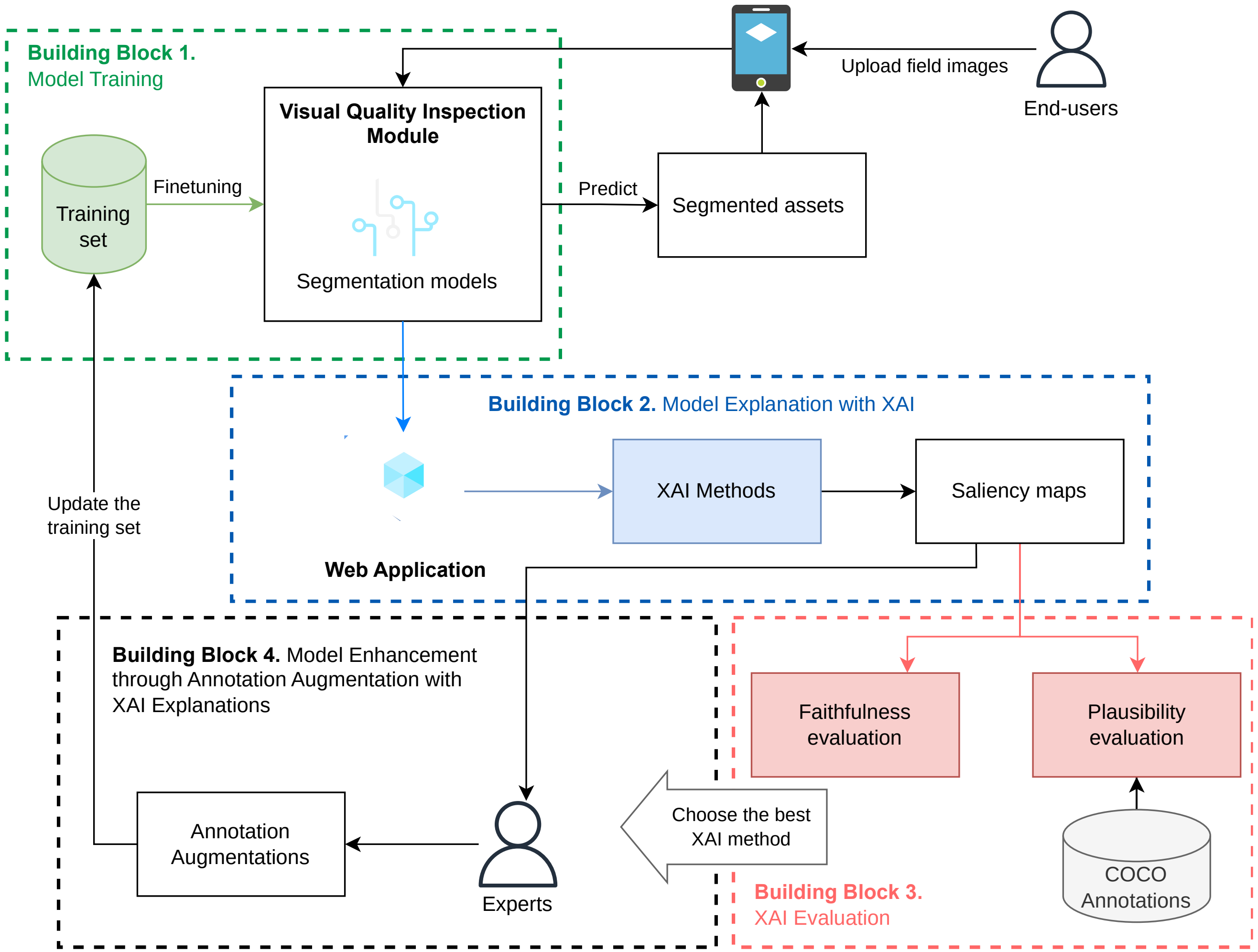
Methodology

Enhanced Visual Quality Inspection

Overall Architecture

Enhanced VQI Framework comprises 4 building blocks:

1. **Model Training:** focuses on training semantic segmentation models.
2. **Model Explanation with XAI:** integrates XAI methods into models to generate explanations for their predictions.
3. **XAI Evaluation:** evaluates the XAI methods using qualitative and quantitative metrics, ensuring the explanations are accurate and understandable.
4. **Model Enhancement with XAI:** enhances the model's performance by augmenting annotations with XAI explanations.



Enhanced Visual Quality Inspection

Dataset – TTPLA (Transmission Towers and Power Lines Aerial-image)

We employ the public TTPLA dataset for segmenting power-grid hardware assets (Abdelfattah et al. 2020).

The dataset comprises 1242 high-resolution images with 8987 instances of transmission towers and power lines, classified into four categories: *cable, tower_wooden, tower_lattice, tower_tucohy.*

The images, annotated in the COCO format, present unique challenges due to the nature of the objects and diverse backgrounds, lighting conditions, and object sizes.

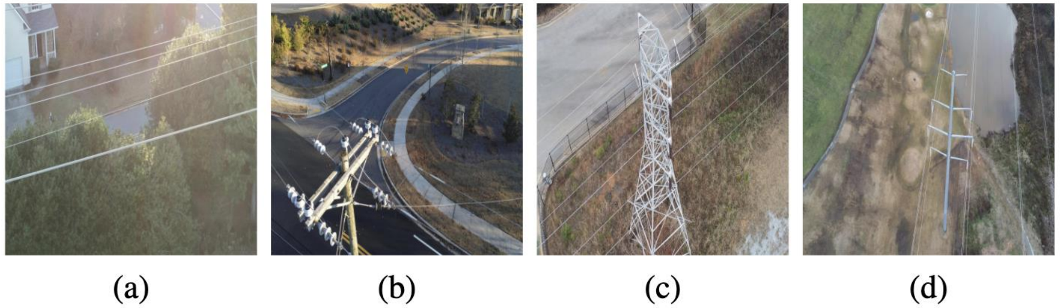


Fig. 1. Data samples from the TTPLA dataset represent the images' main objects of categories (a) cable, (b) tower_wooden, (c) tower_lattice, (d) tower_tucohy.

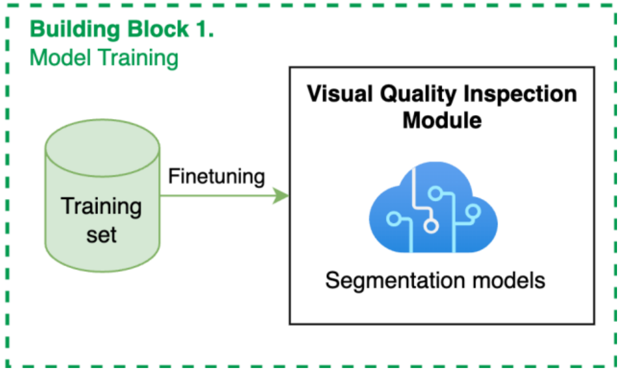
Enhanced Visual Quality Inspection

Building Block 1 – Model Training

This block focuses on the training models: **FCN-VGG16, DeepLabv3-ResNet50, DeepLabv3-ResNet101.**

Splitting dataset: 80%-20% training-test set, with all images resized to 500x500 pixels.

Loss function: Dice loss, which is particularly useful for imbalanced classes in the image segmentation task, as it considers the overlap between the predicted and ground truth masks (Sudre et al. 2017)



Building Block 1: Model Training

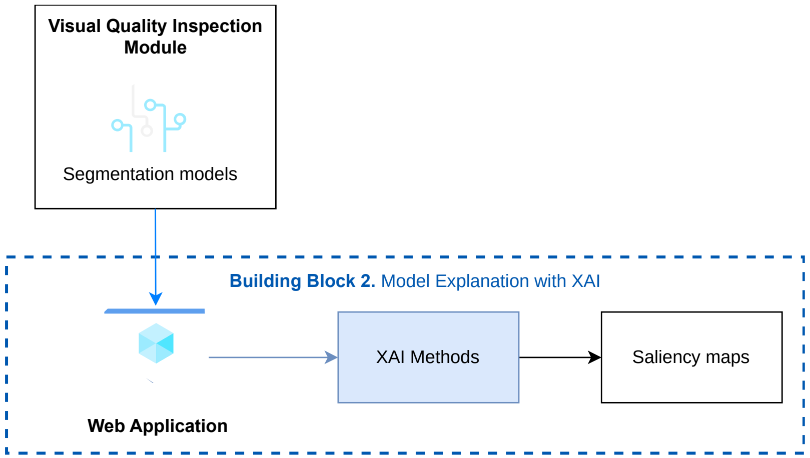
Enhanced Visual Quality Inspection

Building Block 2 – Model Explanation with XAI

Explanation maps of all methods are extracted from the predictions of the segmentation model on the test set, which will be used for the evaluation step.

XAI methods utilized:

- GradCAM (Selvaraju et al. 2017)
- GradCAM++ (Chattopadhay et al. 2018)
- XGradCAM (Fu et al. 2020)
- HiResCAM (Draelos and Carin 2020)
- ScoreCAM (Wang et al. 2020)



Building Block 2: Model Explanation with XAI

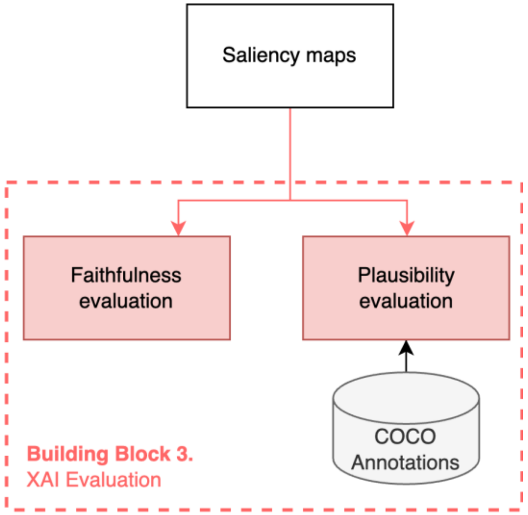
Enhanced Visual Quality Inspection

Building Block 3 – XAI Evaluation

This block evaluates the XAI methods, focusing on the plausibility and faithfulness of their explanations.

Plausibility, the alignment of explanations with human intuition:

- **Energy-Based Pointing Game (EBPG)** (Wang et al. 2020) evaluates the precision and denoising ability of XAI methods to identify the most influential region in an image for a given prediction.
- **Intersection over Union (IoU)** (Bau et al. 2017; Chang et al. 2018) assesses the localization capability and the significance of the attributions captured in an explanation map.
- **Bounding Box (Bbox)** (Schulz et al. 2020) is a variant of the IoU metric that adapts to the size of the object of interest.



Building Block 3: XAI Evaluation

Faithfulness, the alignment of explanations with the model's predictive behavior:

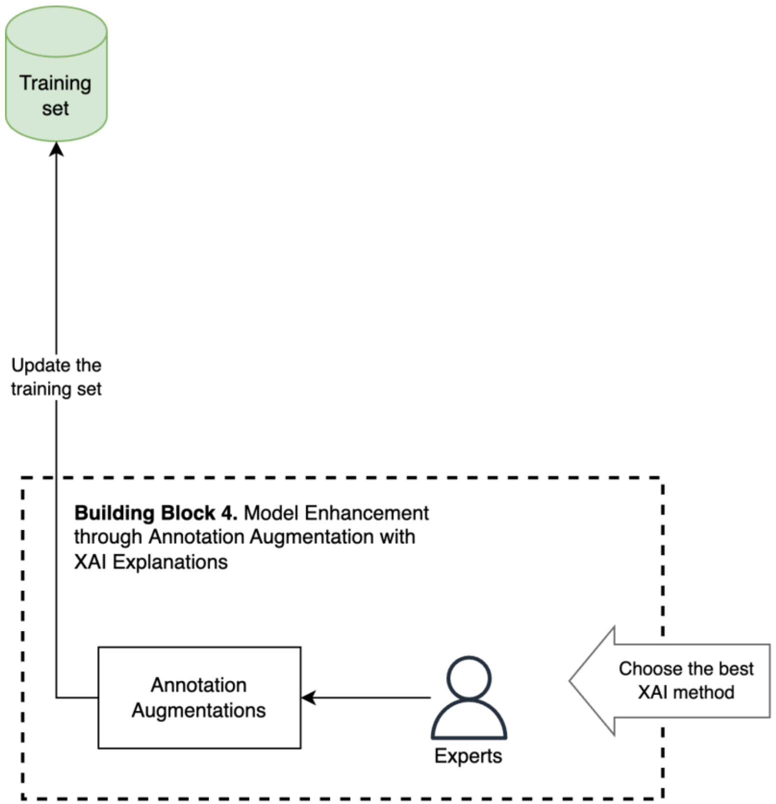
- **Drop** (Fu et al. 2020) measures the average decrease in the model's prediction when the explanation is used as input.
- **Increase** (Fu et al. 2020) quantifies the frequency at which the model's confidence increases when the explanation is used as input.

Enhanced Visual Quality Inspection

Building Block 4 – Model Enhancement with XAI

This block enhances the model's performance based on XAI explanations and expert’s suggestion.

- Data augmentation strategies, such as altering data distribution or adjusting data and labels, have been used to enhance model performance (Zhang et al. 2020, Chu et al. 2021).
- A XAI method is utilized to guide annotation augmentation.
- The COCO annotations from the TTPLA dataset are relabeled based on expert recommendations.
- The model is then retrained on the enhanced training dataset with augmented annotations.



Building Block 4: Model Enhancement with XAI

Results

Building Block 1 – Model Training

Quantitative Result of Model Performance

The performance of the three models is assessed using the IoU metric (Murphy 1996).

DeepLabv3-ResNet101 model demonstrates the best overall performance across the four categories.

Hence, **DeepLabv3-ResNet101** is chosen as the main model in next building blocks, as it has a good performance in the segmentation task and is lightweight enough for the explanation task.

Model	cable	tower_wooden	tower_lattice	tower_tucohy	Overall
DeepLabv3-ResNet50	**56.19**	90.11	92.19	89.68	82.04
DeepLabv3-ResNet101	55.06	**94.75**	**95.31**	**90.63**	**83.94**
FCN-VGG16	52.94	86.49	88.19	73.11	75.18

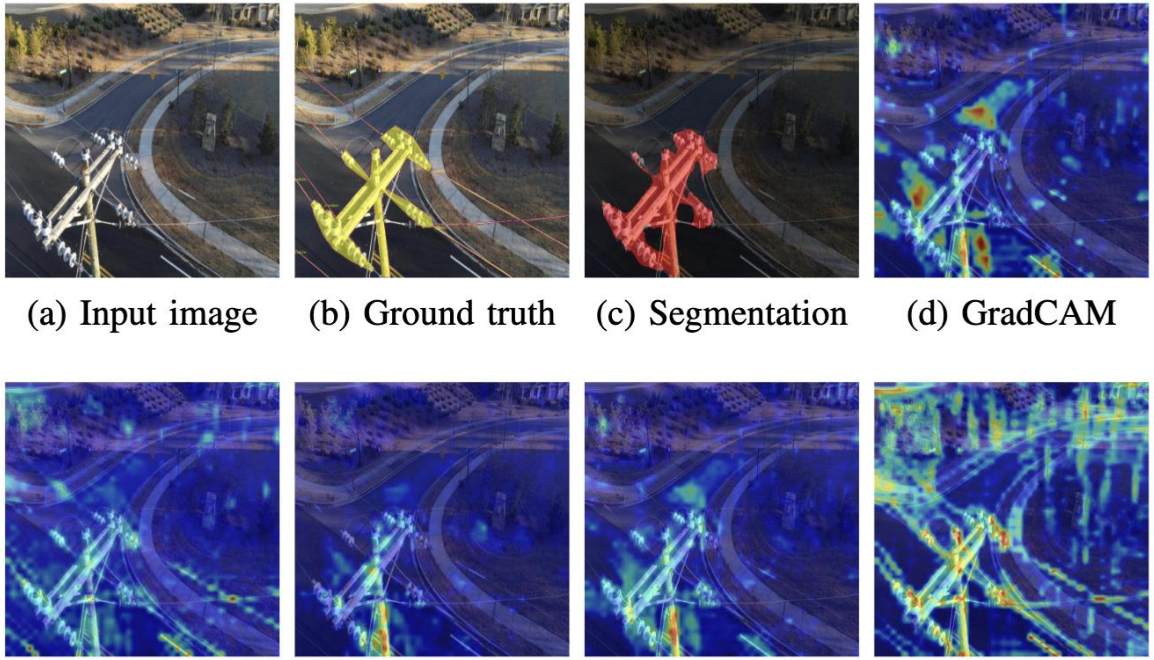
Table 2: The performance of three semantic segmentation models in IoU (%) on each category (cable, tower_wooden, tower_lattice, tower_tucohy) and in average. The best is indicated in bold.

Building Block 2 – Model Explanation with XAI

Qualitative Evaluation

The explanation maps of implemented XAI methods are extracted on the DeepLabv3-ResNet101 on the test set.

Since each method delivers different explanation maps with varied behavior, we need to proceed to Building Block 3 – XAI Evaluation to choose a suitable XAI method for the Model Enhancement step.



(a) Input image (b) Ground truth (c) Segmentation (d) GradCAM

(e) GradCAM++ (f) HiResCAM (g) XGradCAM (h) ScoreCAM

Fig. 3. The qualitative evaluation of implemented XAI methods on the segmentation result of the DeepLabv3-ResNet101 model on a sample from the test set. The category for the segmentation is the tower_wooden denoted under the yellow box shown in the ground truth. The IoU value between the segmentation and the ground truth is 0.9085.

Building Block 3 – XAI Evaluation

Quantitative Evaluation

HiResCAM and GradCAM++ also demonstrate strong performance across several metrics.

Here, we choose HiResCAM as the main XAI method for the next step due to its time-efficiency.

TABLE I

THE QUANTITATIVE EVALUATIONS OF XAI METHODS. FOR EACH METRIC, THE ARROW ↑ / ↓ INDICATES HIGHER/LOWER SCORES ARE BETTER. THE BEST IS IN BOLD.

Method	EPBG ↑	BBox ↑	IoU ↑	Drop ↓	Inc ↑	Time(s) ↓
GradCAM	50.49	48.39	47.94	5.21	52.57	3.21
GradCAM++	58.13	**52.24**	**53.22**	5.17	54.66	4.20
HiResCAM	**60.81**	41.69	52.19	**5.01**	**55.93**	**3.13**
XGradCAM	57.94	47.81	53.09	5.94	55.01	4.43
ScoreCAM	54.01	43.95	51.94	7.34	47.19	52.50

Building Block 4 – Model Enhancement with XAI

Bad performance on the cable class

- The model effectively segments the cable from a clean or mixed-objects background.
- However, when the background contains objects resembling the target object, the model's performance decreases.
- The explanations reveal that the model's attention is directed at the object itself.
- However, in complex cases, the model lacks contextual attention to surrounding objects and background.

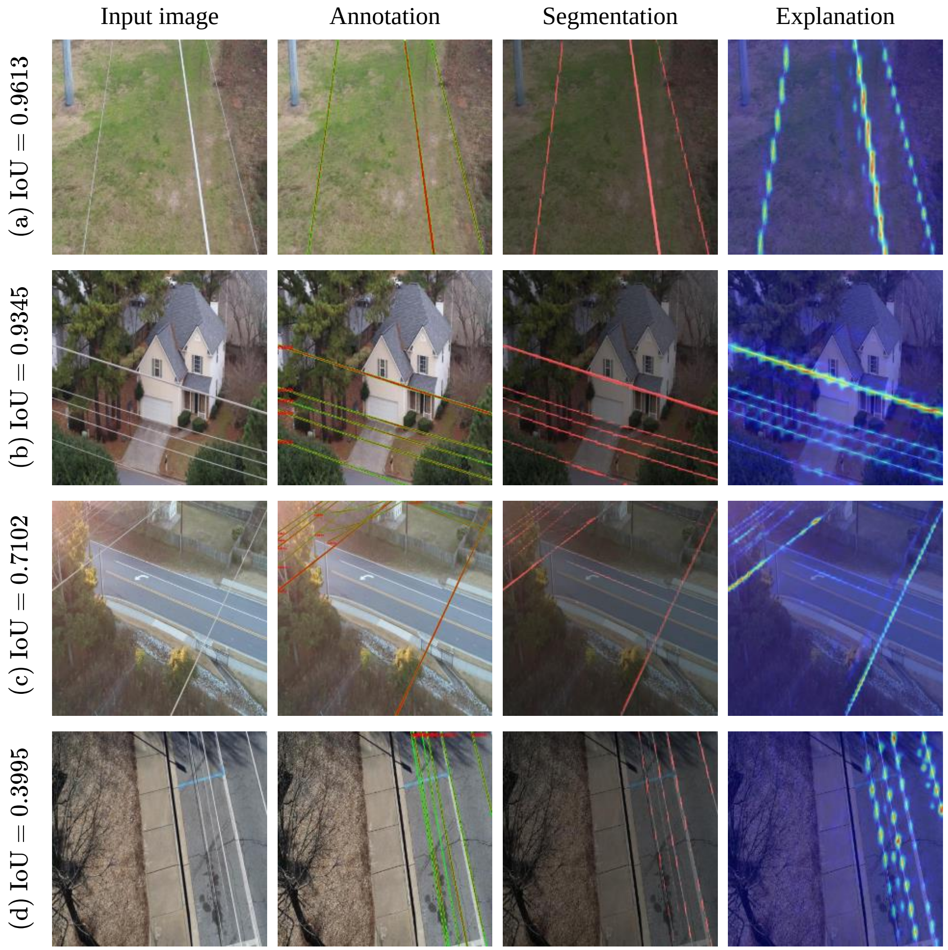


Fig. The list of input images, COCO annotations (ground truth), segmentation results of the DeepLabv3-ResNet101 model, and the explanations in the increasing order of complexity.

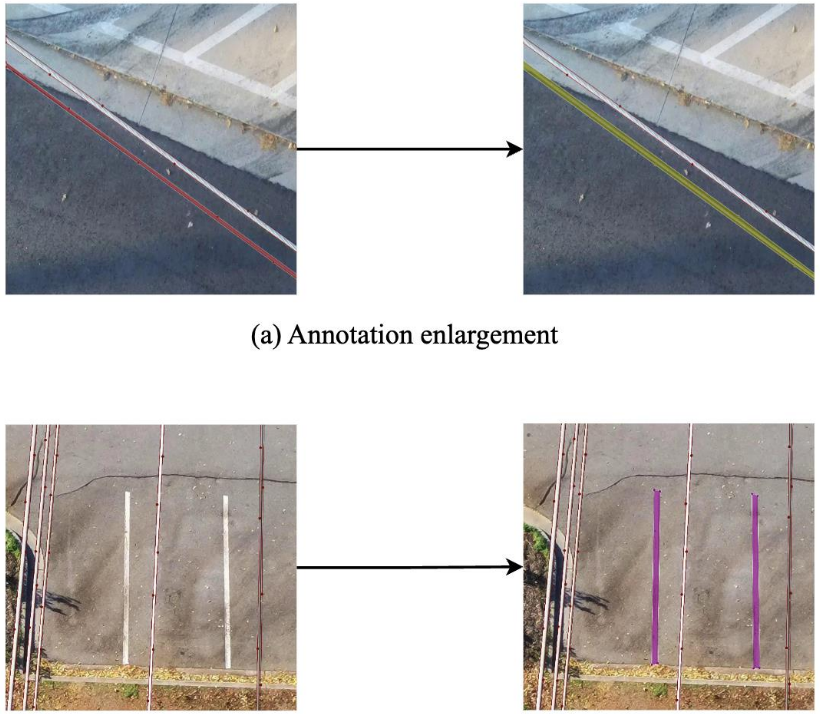
Building Block 4 – Model Enhancement with XAI

Approaches

Therefore, as we employ the annotation augmentation procedure to enhance the model's performance, a domain expert is enlisted to suggest 2 annotation augmentation approaches for each sample:

- **Annotation enlargement:** Given that the model can leverage surrounding contextual information to improve its performance, we propose to enlarge the annotations of thin objects, especially thin cables, which the model often overlooks based on the saliency maps. We increase the object's size by 2 pixels on both sides.

- **Adding annotations for perplexed objects:** As the model often confuses cables with perplexing objects like road surface markings, we propose adding `void` annotations to categorize these perplexing objects as unlabeled objects.



(a) Annotation enlargement

(b) Adding annotations for perplexed objects

Figure 6: Annotation augmentation approaches: (a) Annotation enlargement where the size of the annotation for thin objects like cables is increased, (b) Adding annotations for perplexed objects like the road surface marks to guide the model in differentiating between white cables and perplexed objects.

Building Block 4 – Model Enhancement with XAI

Enhancement Results

Fig. 6. Qualitative results of DeepLabv3-ResNet101 before and after applying the enhancing model performance by annotation augmentation with XAI methods procedure.

Model	cable	tower_wooden	tower_lattice	tower_tucohy	Overall
DeepLabv3-ResNet101	55.06	94.75	95.31	90.63	83.94
DeepLabv3-ResNet101 (Enhanced)	**58.11**	**94.78**	**95.32**	**90.65**	**84.715**

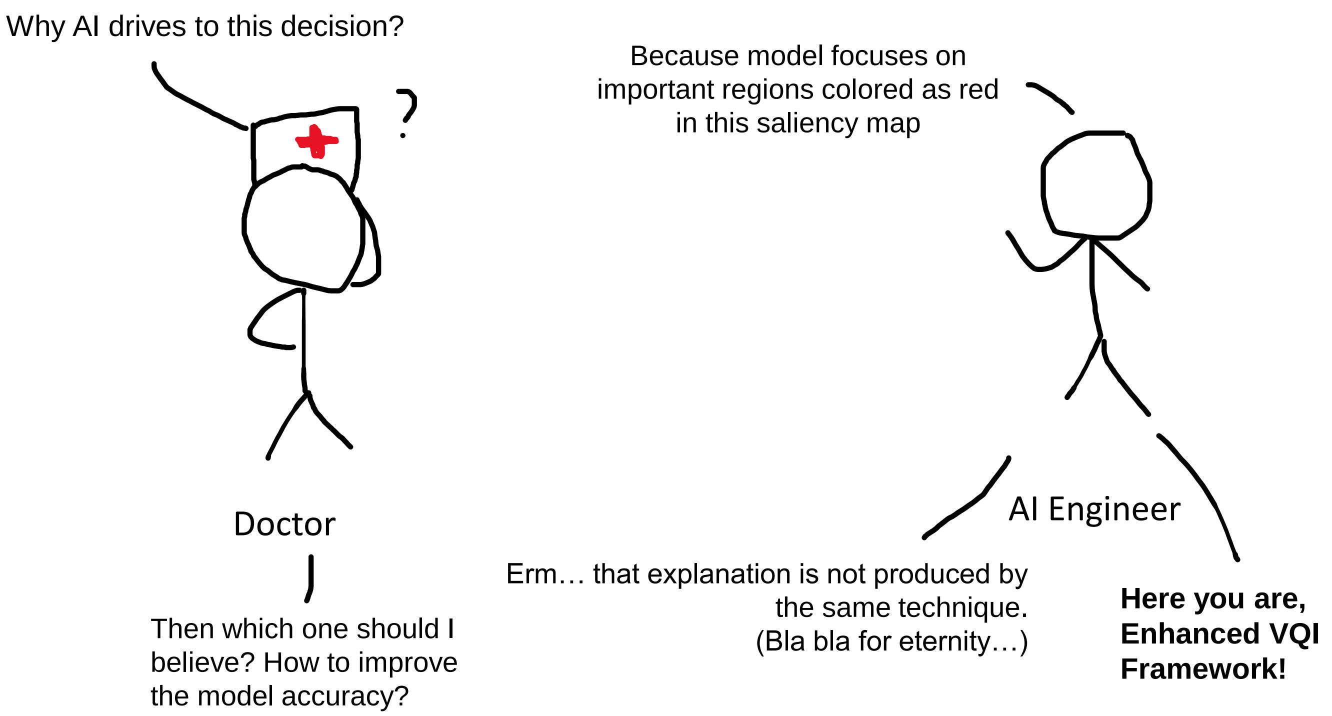
Table 4: The quantitative results of semantic segmentation models DeepLabv3-ResNet101 before and after applying the enhancing model performance by annotation augmentation with XAI methods procedure in IoU (%) on each category (`cable`, `tower_wooden`, `tower_lattice`, `tower_tucohy`) and in average. The better is indicated in bold.

Conclusion & Future Work

Conclusion

This work proposed an Enhanced VQI Framework employing XAI techniques to enhance interpretability and performance in semantic segmentation tasks.

The model enhancement procedure, guided by XAI's explanation maps, effectively improved model performance in complex object segmentation and detection, especially in challenging contexts where objects and backgrounds are indistinguishable.



Future of XAI

BEFORE

Training data → ML process → Learned Function → Decision Cat: 0.76 → End-user

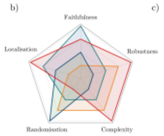
- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

TODAY

Training data → New ML process → (Explainable) Learned Func. | Explainable decision ⇄ End-user

- I understand why
- I understand why not
- I know when you succeed
- I know when you fail
- I know when to trust you
- I know why you erred

TOMORROW



Evaluation

Training data → New ML process → (Explainable) Learned Func. | Explainable decision ⇄ End-user

Feedback

- Which explanation is better?
- How does this explanation mean?
- Can I trust the explanation?
- How to use XAI to improve models?