



Multimodal Explanations for AI- based Multisensor Fusion

***Dave Braines** (IBM Research, UK), **Alun Preece**, **Dan Harborne** (Crime & Security Research Institute, Cardiff University, UK)*

[dstl]

ARL

Introduction & scope

AI techniques can be inscrutable

- We don't know "why?"...

Lots of active research into XAI

- Explainable Artificial Intelligence (XAI)
- Multi-modality data and explanation

We have developed a conceptual model

- To underpin explanations
- For human and machine processing

A conversational interface

- To explore explanations

Worked examples and an everyday scenario



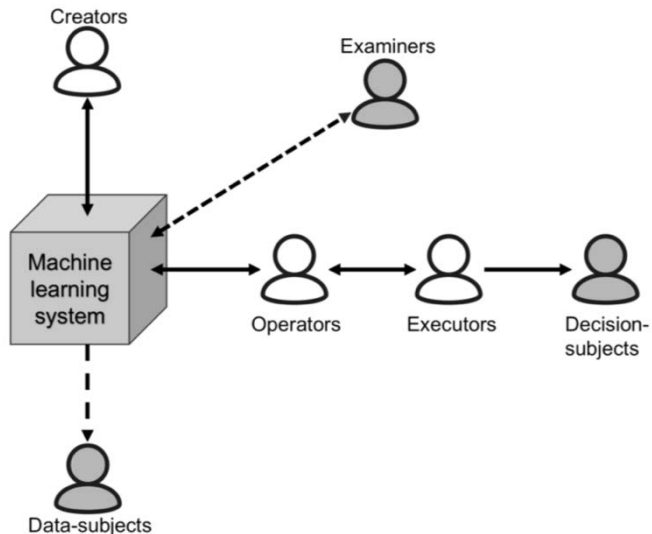
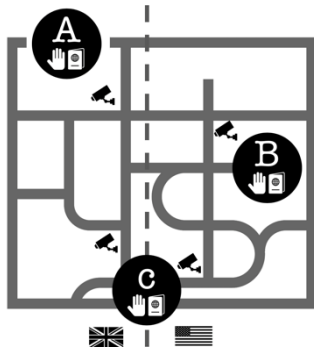
Background

Scenario and dataset

- Real-time London CCTV imagery
- Coalition context & edge processing
- Many derivative datasets possible

Explanation-oriented architecture (XOA)

- Rapid ensemble services
- Trust and confidence



Explanation types

- Transparent, post-hoc
- Multiple modalities

Conversation and roles

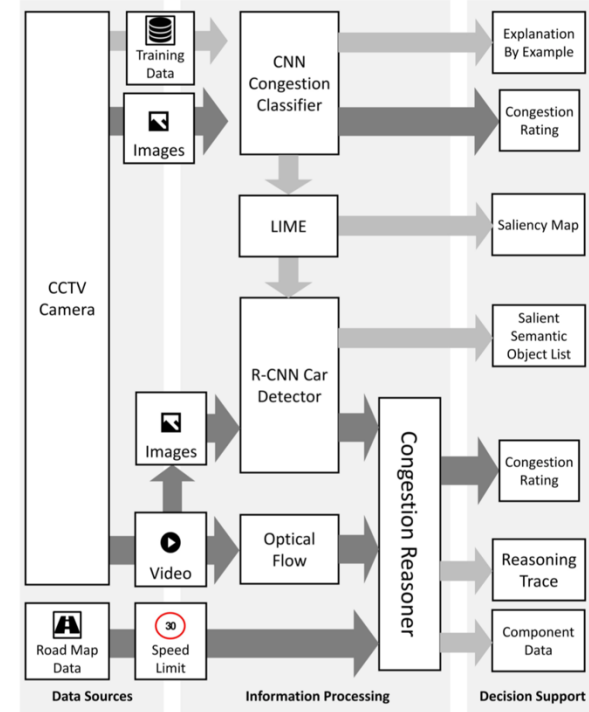
- We treat explanation as a conversation
- User role and task context are key

Worked Example

Using our Explanation Oriented Architecture

- Detect or infer traffic congestion
- Congestion & explanation services and flows
- Information fusion from multi-modal data sources

Three types of congestion services:



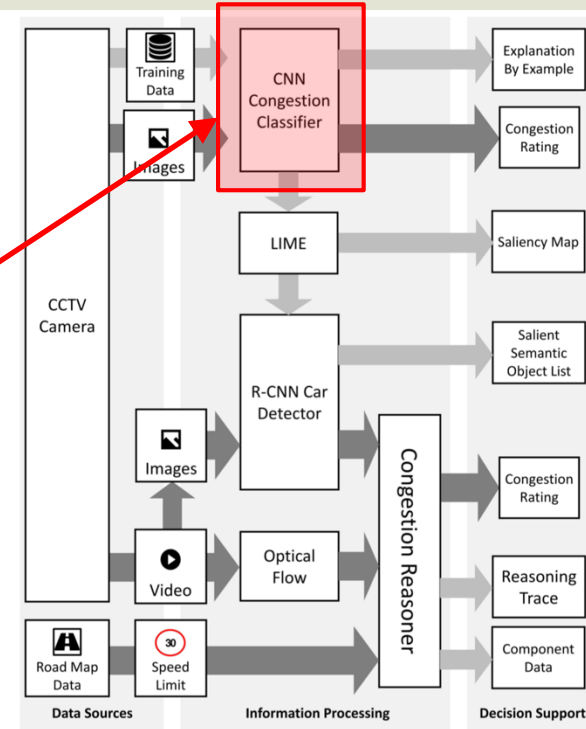
Worked Example

Using our Explanation Oriented Architecture

- Detect or infer traffic congestion
- Congestion & explanation services and flows
- Information fusion from multi-modal data sources

Three types of congestion services:

1. Congestion Image Classifier (CIC)



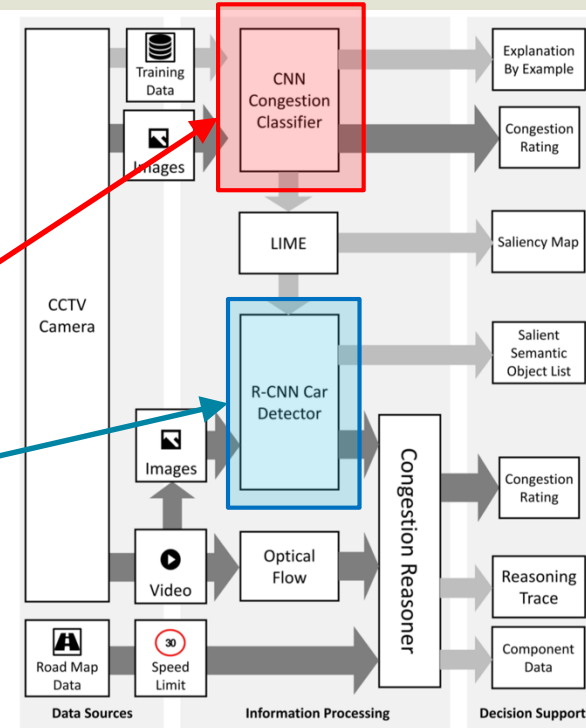
Worked Example

Using our Explanation Oriented Architecture

- Detect or infer traffic congestion
- Congestion & explanation services and flows
- Information fusion from multi-modal data sources

Three types of congestion services:

1. Congestion Image Classifier (CIC)
2. Entity detector (ED)



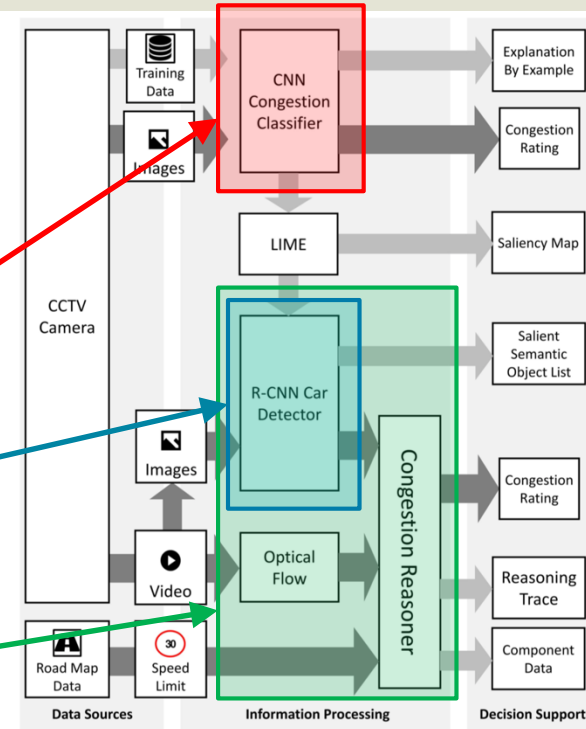
Worked Example

Using our Explanation Oriented Architecture

- Detect or infer traffic congestion
- Congestion & explanation services and flows
- Information fusion from multi-modal data sources

Three types of congestion services:

1. Congestion Image Classifier (CIC)
2. Entity detector (ED)
3. Congestion Speed Classifier (CSC)



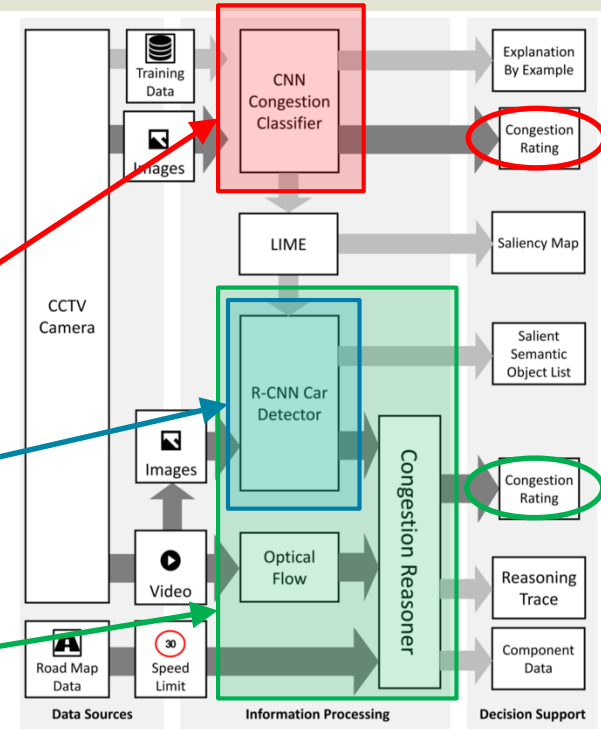
Worked Example

Using our Explanation Oriented Architecture

- Detect or infer traffic congestion
- Congestion & explanation services and flows
- Information fusion from multi-modal data sources

Three types of congestion services:

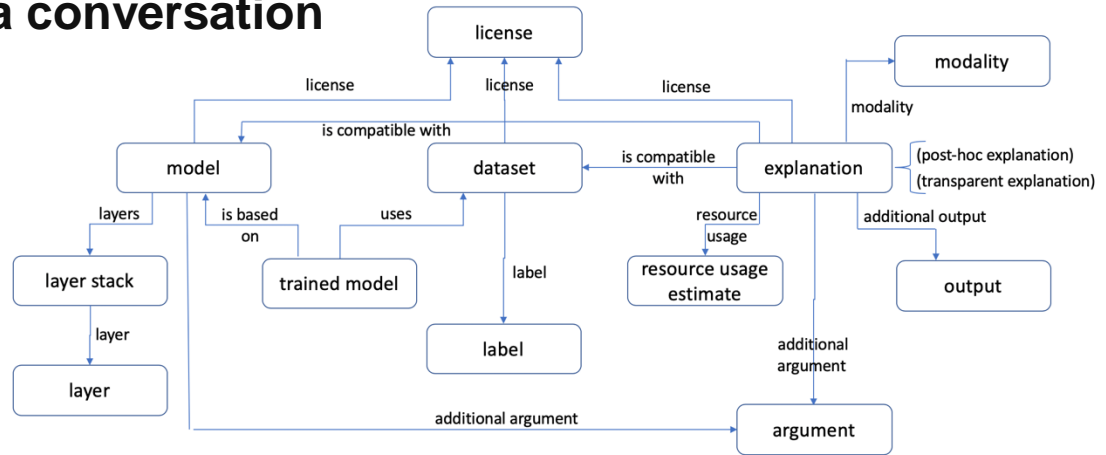
1. Congestion Image Classifier (CIC)
2. Entity detector (ED)
3. Congestion Speed Classifier (CSC)



Conversations for Explanation

Explanation takes the form of a conversation

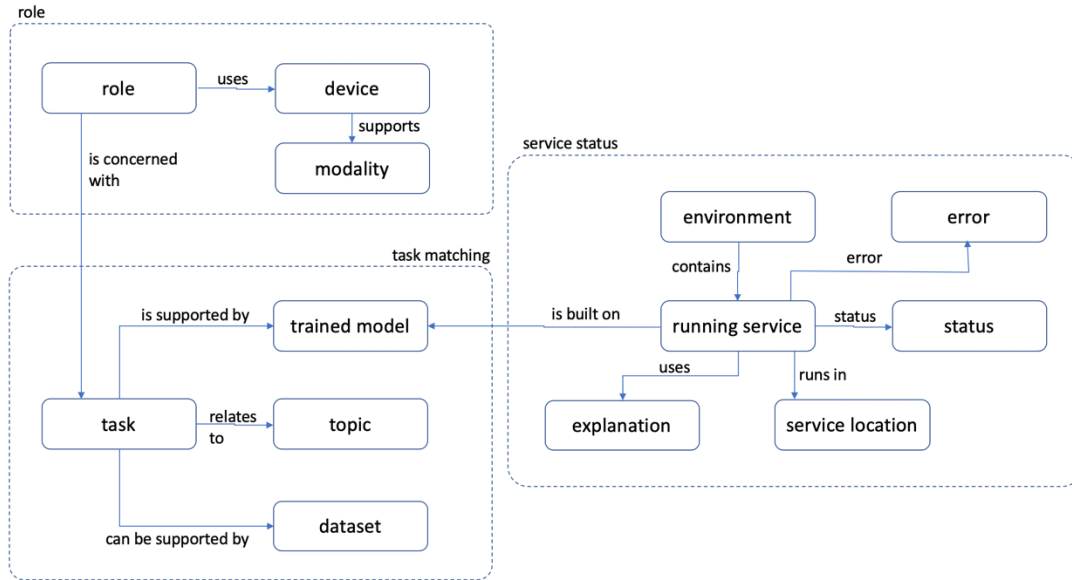
- Can be in many forms
- Focus on the abstract space
- Text-chat format
- Embedded additional modalities



Conceptual model defines key concepts and relationships

- Ontology, but for machine and human processing
- Key explanation-related concepts in “core” model

Conversations for Explanation



Additional models easily integrated

- To define the context for the explanations
- Role, task, services, etc

*The **ad-hoc** and **real-time** operational environment require **agile context modelling**.*



Conversations for Explanation

Use case 1: Fully transparent



Conversations for Explanation

Use case 1: Fully transparent

Is there congestion at Checkpoint A?



Conversations for Explanation

Use case 1: Fully transparent

Is there congestion at Checkpoint A?

No, I am confident that checkpoint A is not congested.



Conversations for Explanation

Use case 1: Fully transparent

Is there congestion at Checkpoint A?

No, I am confident that checkpoint A is not congested.

Why?



Conversations for Explanation

Use case 1: Fully transparent

Is there congestion at Checkpoint A?

No, I am confident that checkpoint A is not congested.

Why?

Live CCTV shows 12 moving objects at 80% of the speed limit which indicates it is not congested.

Conversations for Explanation

Use case 2a: Post-hoc by Saliency



Conversations for Explanation

Use case 2a: Post-hoc by Saliency

Is there congestion at Checkpoint B?



Conversations for Explanation

Use case 2a: Post-hoc by Saliency

Is there congestion at Checkpoint B?

I am not confident, but Checkpoint B is congested.



Conversations for Explanation

Use case 2a: Post-hoc by Saliency

Is there congestion at Checkpoint B?

I am not confident, but Checkpoint B is congested.

Why?



Conversations for Explanation

Use case 2a: Post-hoc by Saliency

Is there congestion at Checkpoint B?

I am not confident, but Checkpoint B is congested.

Why?

This live CCTV image shows the situation and I believe Checkpoint B is congested (with 62% confidence).



The green areas show the parts of the image that most indicate congestion.

Conversations for Explanation

Use case 2b: Post-hoc by Example



Conversations for Explanation

Use case 2b: Post-hoc by Example

Is there congestion at Checkpoint C?



Conversations for Explanation

Use case 2b: Post-hoc by Example

Is there congestion at Checkpoint C?

I am not confident, but Checkpoint C is congested.



Conversations for Explanation

Use case 2b: Post-hoc by Example

Is there congestion at Checkpoint C?

I am not confident, but Checkpoint C is congested.

Why?



Conversations for Explanation

Use case 2b: Post-hoc by Example

Is there congestion at Checkpoint C?

I am not confident, but Checkpoint C is congested.

Why?

I can't show you the live CCTV image but these other images are rated as being similarly congested.



Conversations for Explanation

Use case 3: Disagreement within services



Conversations for Explanation

Use case 3: Disagreement within services

Is there congestion at Checkpoint A?



Conversations for Explanation

Use case 3: Disagreement within services

Is there congestion at Checkpoint A?

I cannot be confident either way, sorry.



Conversations for Explanation

Use case 3: Disagreement within services

Is there congestion at Checkpoint A?

I cannot be confident either way, sorry.

Why?



Conversations for Explanation

Use case 3: Disagreement within services

Is there congestion at Checkpoint A?

I cannot be confident either way, sorry.

Why?

Live CCTV shows 2 moving objects at 80% of the speed limit which indicates it is not congested. But the live CCTV image is classified as congested.



These outcomes are inconsistent.

The green areas show the parts of the image that most indicate congestion.

Related work

- **Insight from Social Sciences**

Miller, T. (2017). Explanation in artificial intelligence: insights from the social sciences. arXiv preprint arXiv:1706.07269.

- **A grammar for the development of conversational explanations?**

Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., & Mordvintsev, A. (2018). The building blocks of interpretability. Distill, 3(3), e10.

- **Affordances – the strengths of human and machine agents**

Crouser, R. J., & Chang, R. (2012). An affordance-based framework for human computation and human-computer collaboration. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2859-2868.

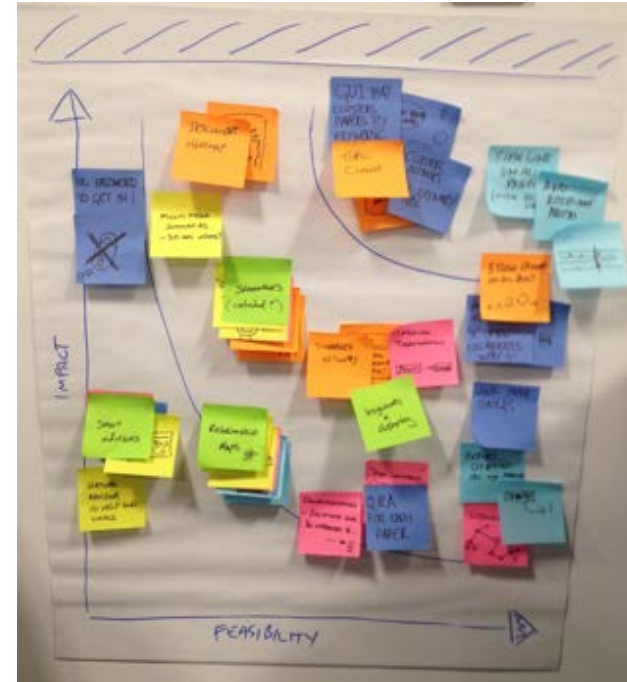
- **Human-Computer Collaboration to drive our conversational principles**

L. Terveen, “Overview of human-computer collaboration,” Knowledge Based Systems, vol. 8(2), pp. 67–81, 1995.



Conclusion & Future work

- **Conversational interface**
 - To explore explanations
 - Human/machine hybrid system
- **Explanations for AI services**
 - Multi-modal data and explanations
 - Rapidly assembled ensembles
 - Coalition context
- **Conceptual model to underpin our approach**
- **Real-world scenario with 3 examples**
- **More interaction with Subject Matter Experts**
- **Experimental design and execution**



Acknowledgment

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

