THALES

On the Role of Domain Knowledge in Explainable Machine Learning

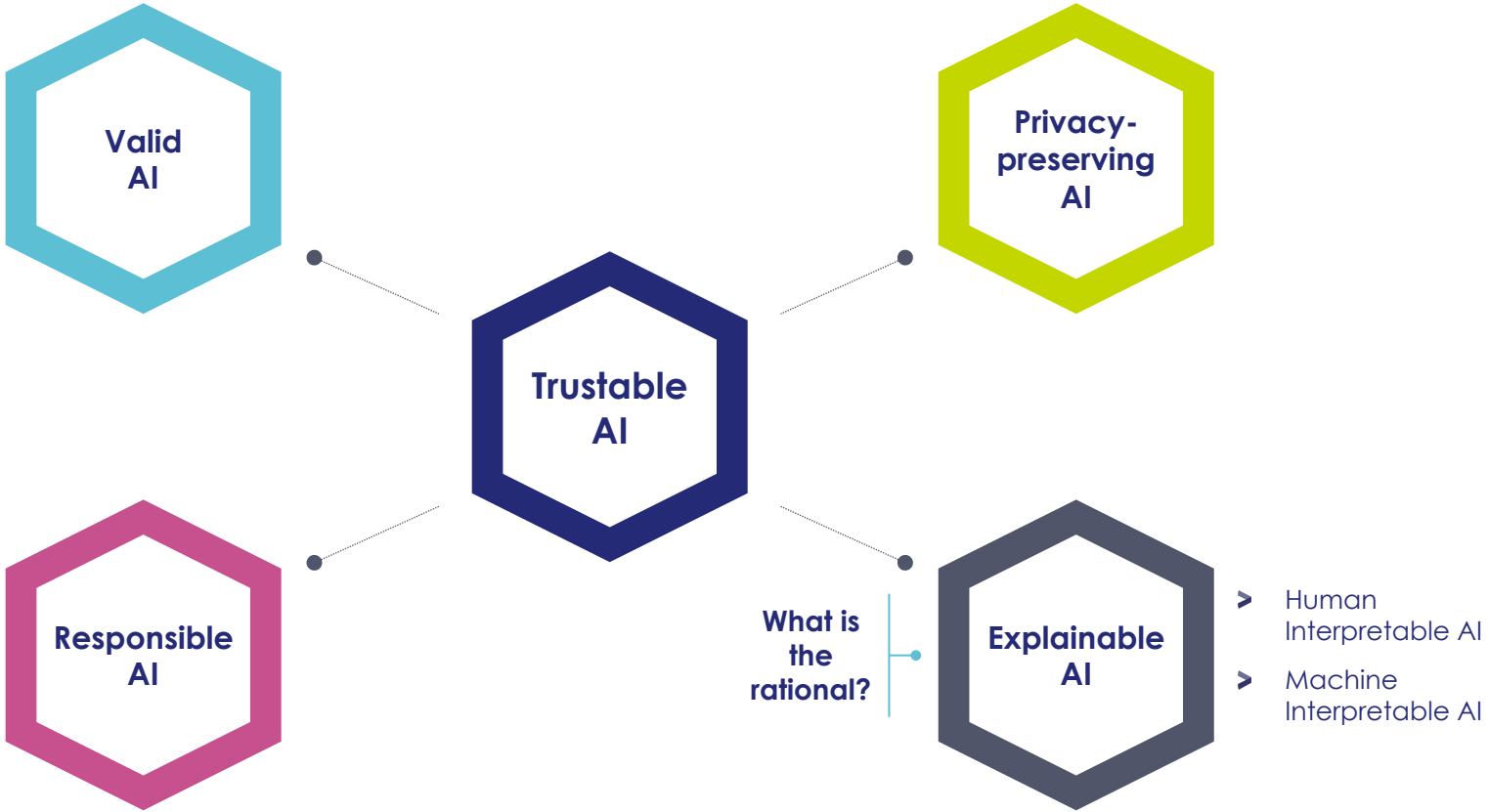
Journées de Statistique de la SFdS 2021
(Virtual Event)
June 11, 2021

Freddy Lecue, Chief AI Scientist (Thales Canada, Inria France)
http://www-sop.inria.fr/members/Freddy.Lecue/
@freddylecue

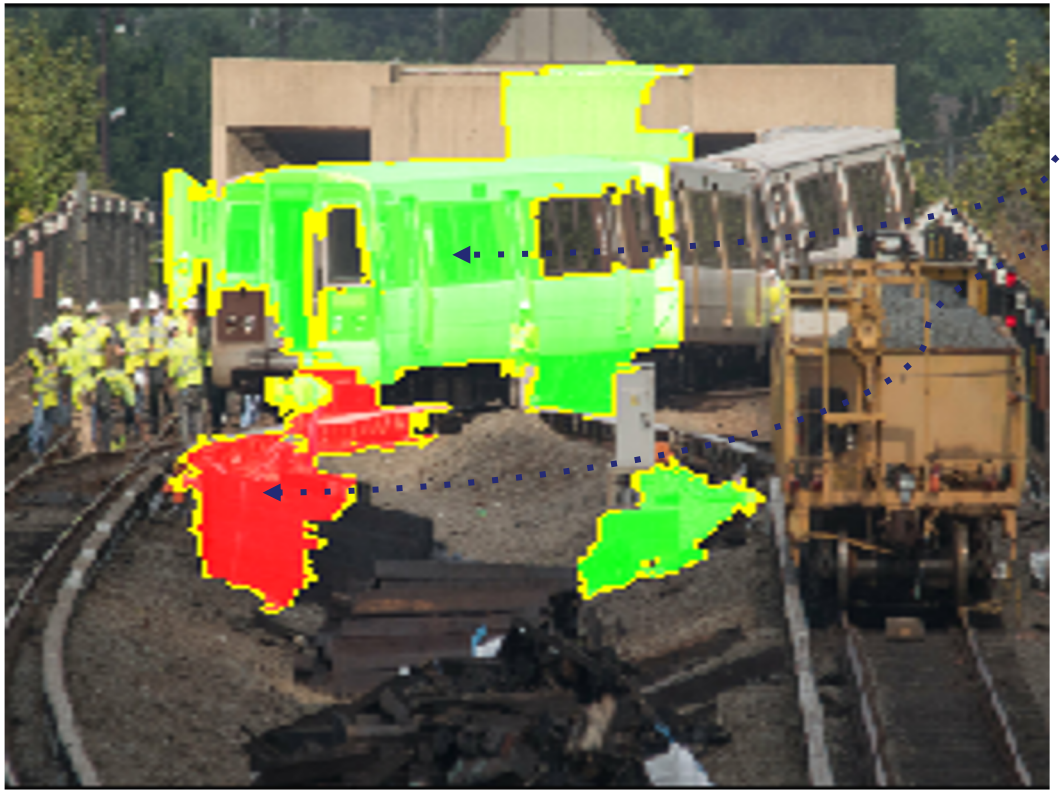
Critical Applications


BNSF

AI Adoption: Requirements



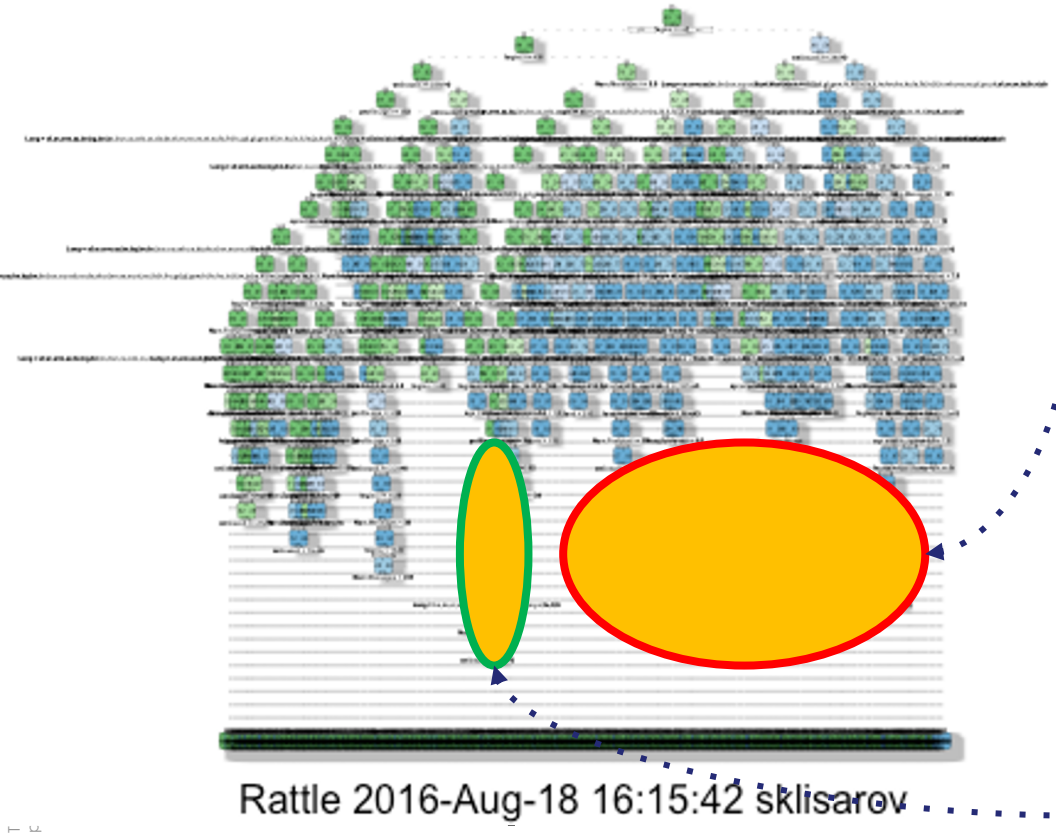
Knowledge Graph in Machine Learning (1)



Augmenting (input) features with more semantics such as knowledge graph embeddings / entities

https://stats.stackexchange.com/questions/230581/decision-tree-too-large-to-interpret

Knowledge Graph in Machine Learning (2)



https://stats.stackexchange.com/questions/230581/decision-tree-too-large-to-interpret

Augmenting machine learning models with more semantics such as knowledge graphs entities

Knowledge Graph in Machine Learning (3)

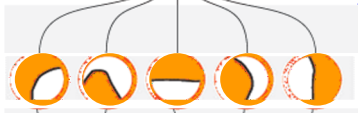
Input Layer

Training Data

Input (unlabeled image)

Hidden Layer

Neurons respond to simple shapes



1st Layer

Neurons respond to more complex structures

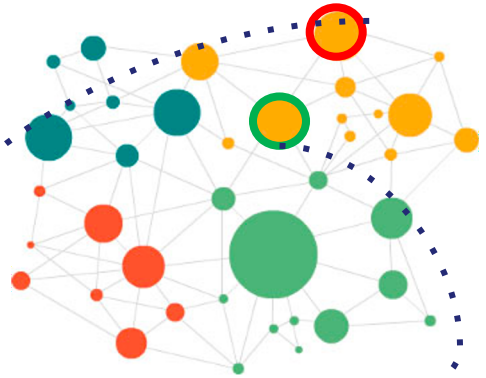
2nd Layer

Neurons respond to highly complex, abstract concepts

nth Layer

Output Layer

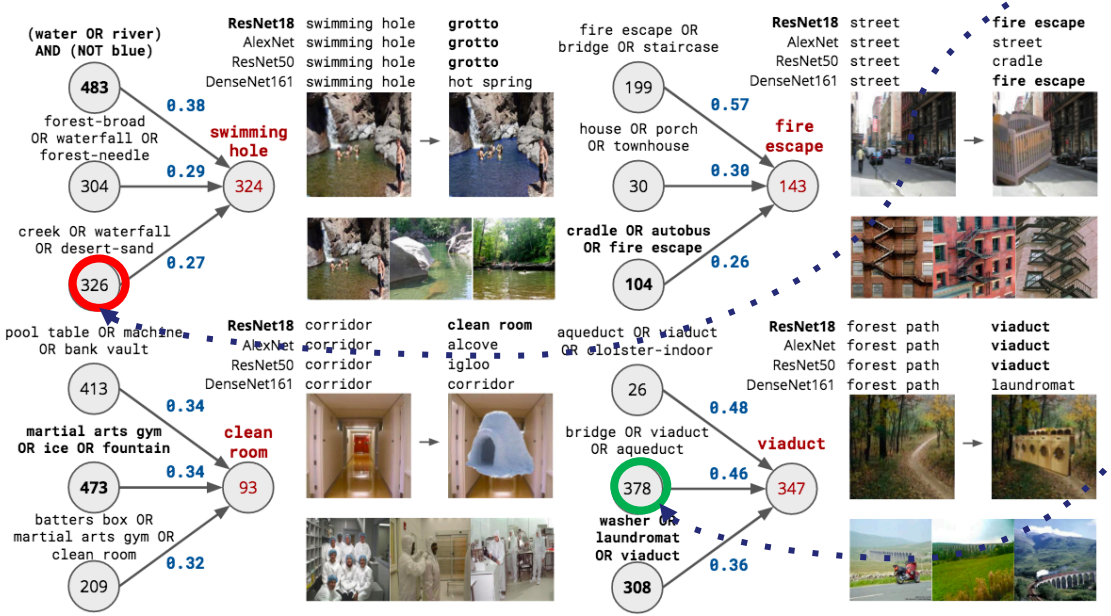
Low-level features to high-level features



Augmenting (intermediate) features with more semantics such as knowledge graph embeddings / entities

THALES

Knowledge Graph in Machine Learning (4)

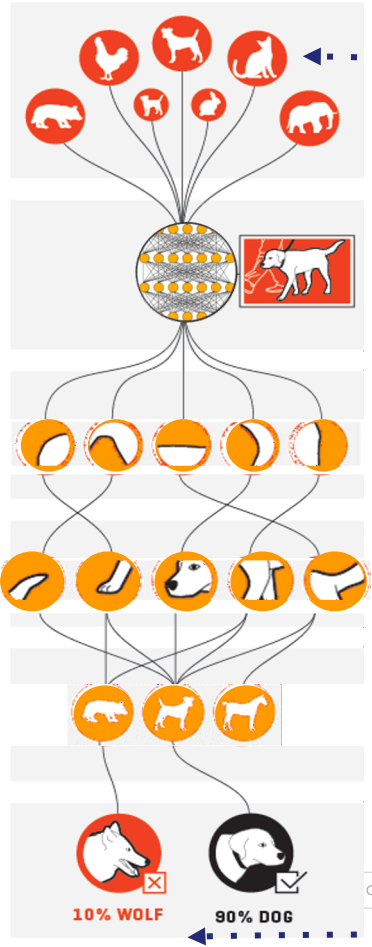


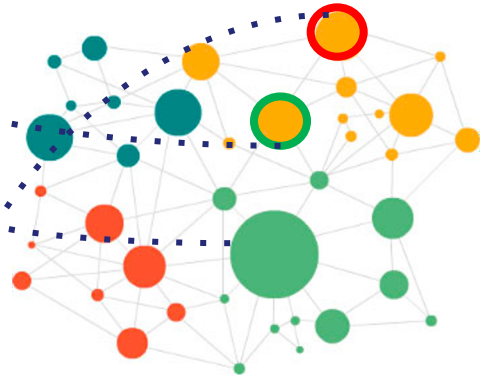
Jesse Mu, Jacob Andreas:Compositional Explanations of Neurons. NeurIPS 2020

Low-level features to high-level features

Open question: What is the impact of semantic representation on units in Neural Networks?

Knowledge Graph in Machine Learning (5)





Low-level features to high-level features

Augmenting (input, intermediate) features – output relationship with more semantics to capture causal relationship

THALES

Knowledge Graph in Machine Learning (6)

Description 1: This is an orange train accident

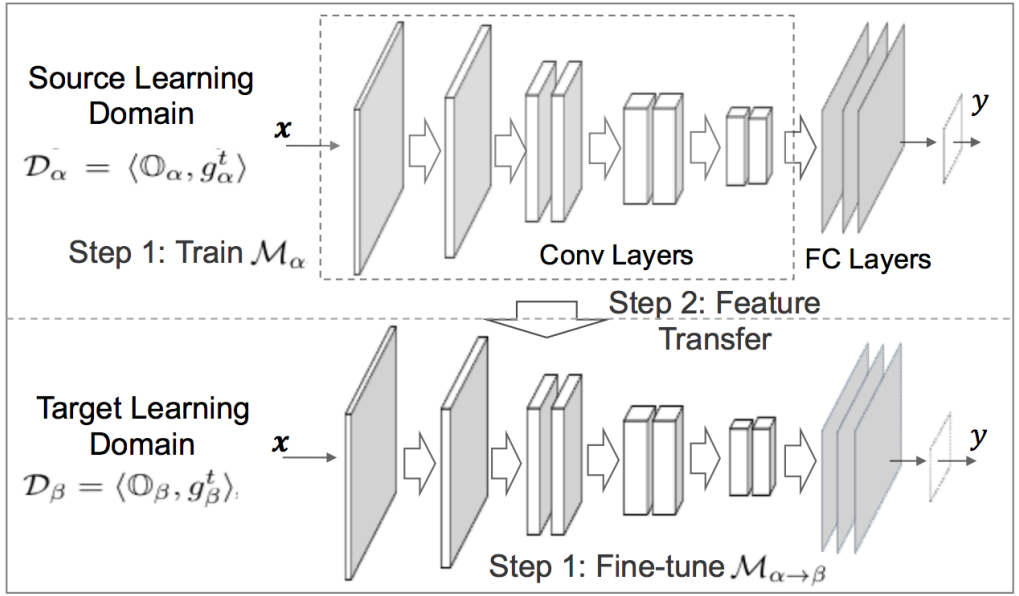
Description 2: This is a train accident between two speed merchant trains of characteristics X43-B and Y33-C in a dry environment

Description 3: This is a public transportation accident

Augmenting models with semantics to support personalized explanation

Knowledge Graph in Machine Learning (7)

"How to explain transfer learning with appropriate knowledge representation?



Augmenting input features and domains with semantics to support interpretable transfer learning

Jiaoyan Chen, Freddy Lécué, Jeff Z. Pan, Ian Horrocks, Huajun Chen: Knowledge-Based Transfer Learning Explanation. KR 2018: 349-358

THALES

Knowledge Graph in Machine Learning (8)

"How to explain concept drift in Machine Learning?

Augmenting input features and domains with semantics to interpret concept drift in Machine Learning

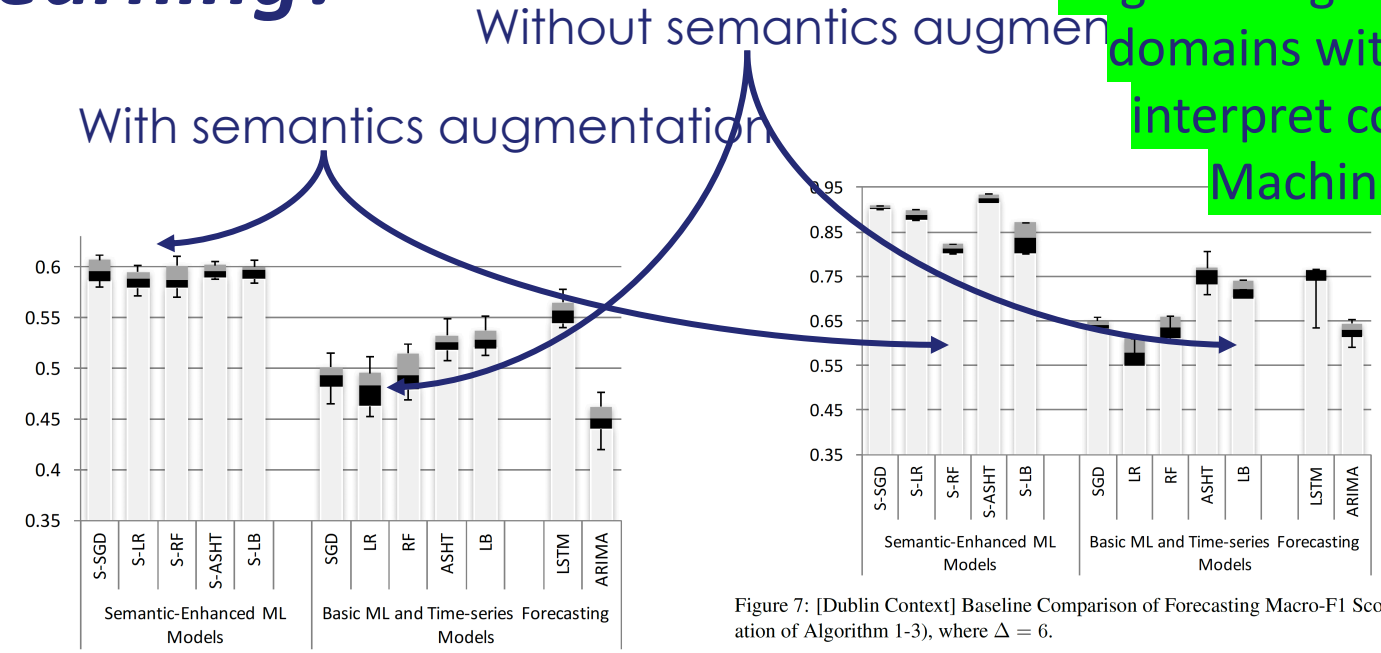


Figure 6: [Beijing Context] Baseline Comparison of Forecasting Macro-F1 Score (Evaluation of Algorithm 1-3), where $\Delta = 6$.

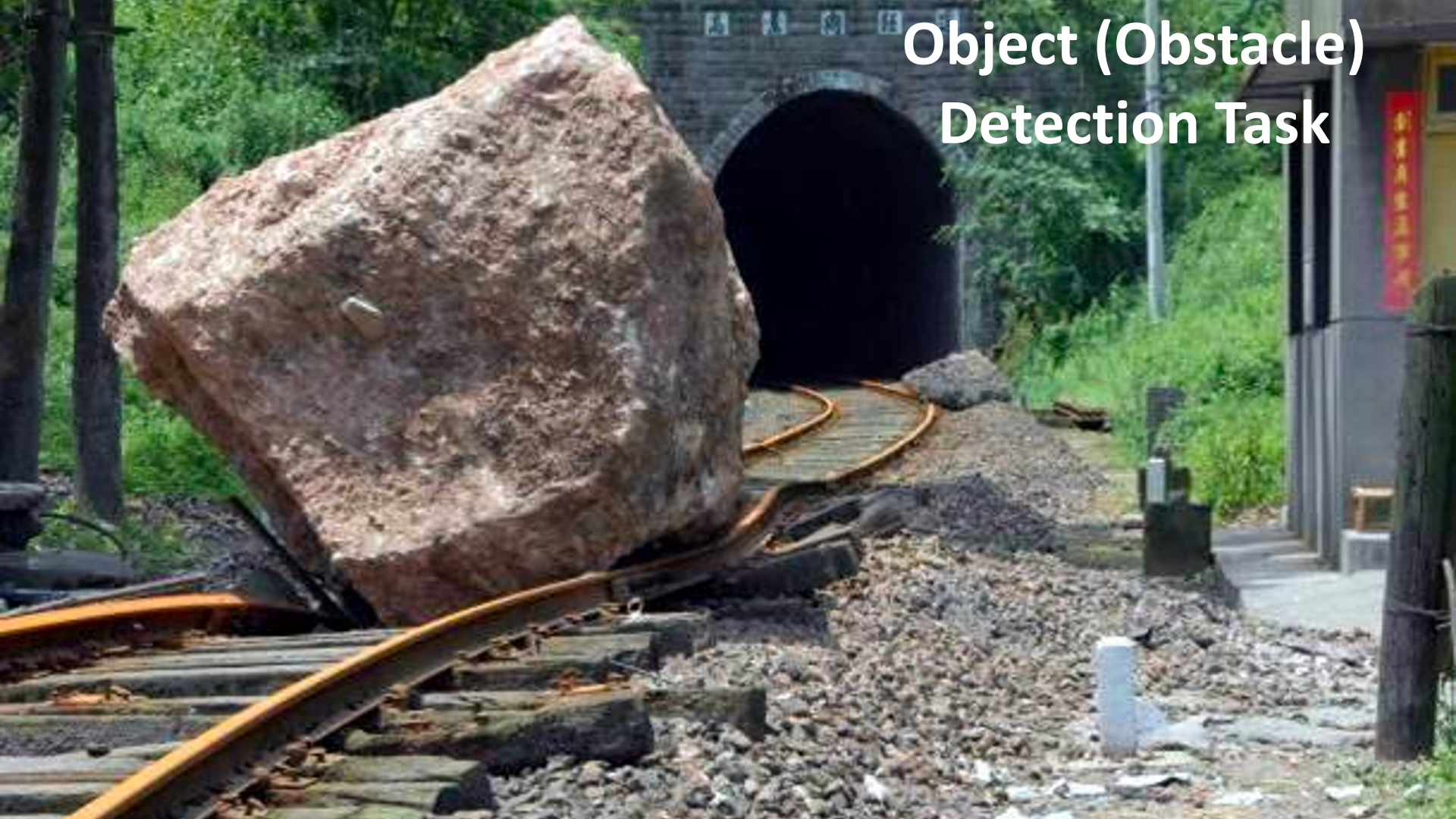
Figure 7: [Dublin Context] Baseline Comparison of Forecasting Macro-F1 Score (Evaluation of Algorithm 1-3), where $\Delta = 6$.

Jiaoyan Chen and Freddy Lécué and Jeff Z. Pan and Shumin Deng and Huajun Chen. Knowledge graph embeddings for dealing with concept drift in machine learning. Journal of Web Semantics. (2021) http://www.sciencedirect.com/science/article/pii/S1570826820300585

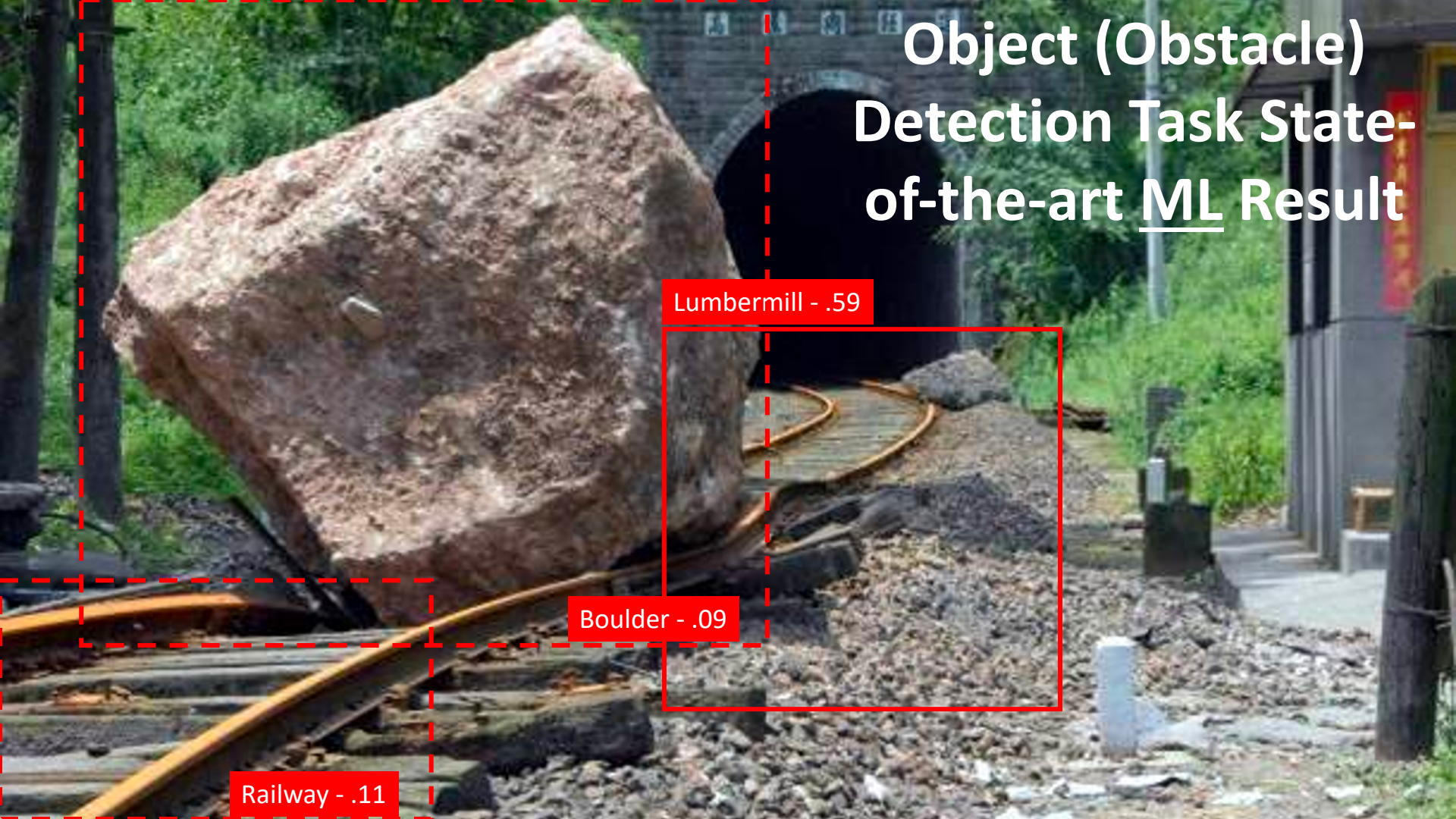
THALES

How Does it Work in Practice?

State of the Art Machine Learning Applied to Critical Systems


Object (Obstacle)
Detection Task

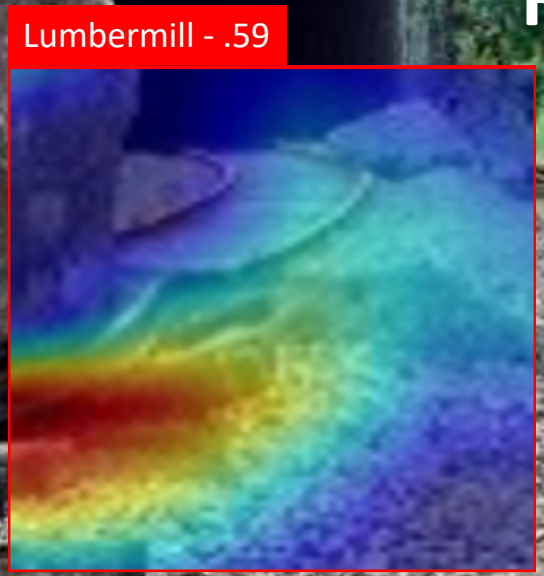
Object (Obstacle) Detection Task State-of-the-art ML Result


Object (Obstacle) Detection Task State-of-the-art ML Result
Lumbermill - .59
Boulder - .09
Railway - .11

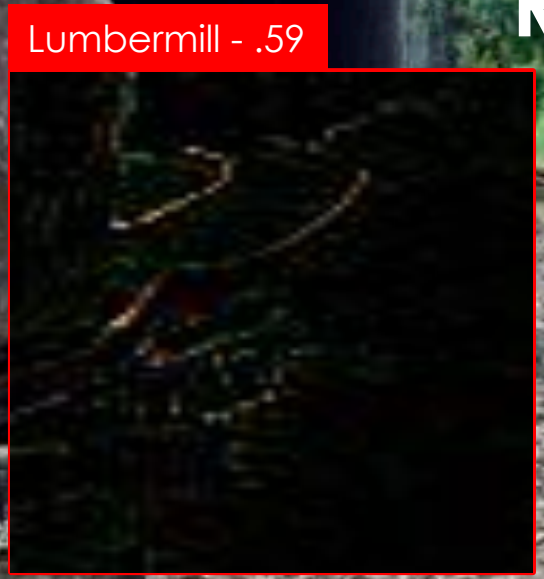
State of the Art XAI Applied to Critical Systems

THALES

Object (Obstacle) Detection Task State-of-the-art XAI Result



Object (Obstacle) Detection Task State-of-the-art XAI Result



Object (Obstacle) Detection Task State-of-the-art XAI Result

Lumbermill - .59

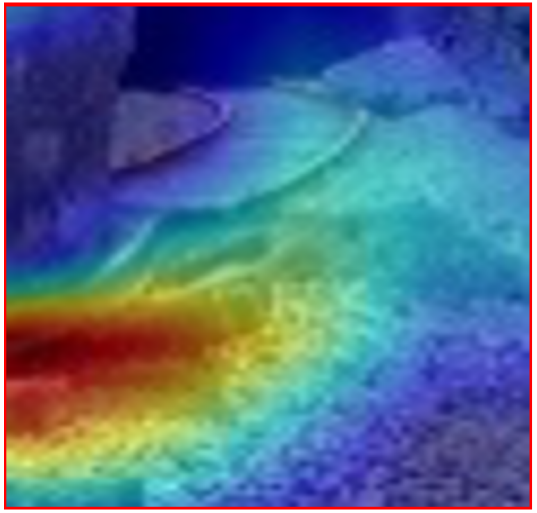
Unfortunately, this is of NO use for a human behind the system

Let’s stay back

Why this Explanation? (meta explanation)

After Human Reasoning…

Lumbermill - .59



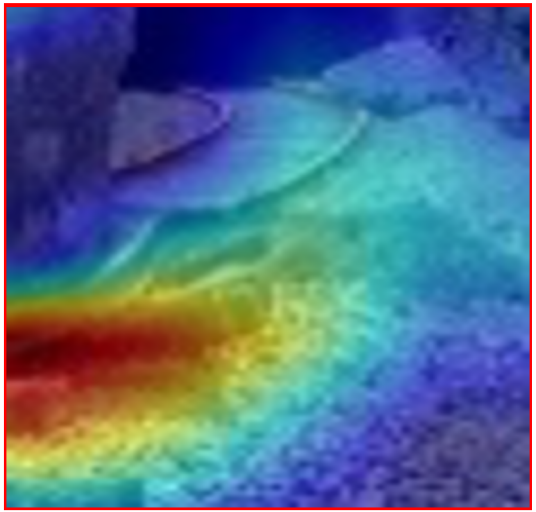
DBpedia Browse using Formats Faceted Browser Sparql Endpoint

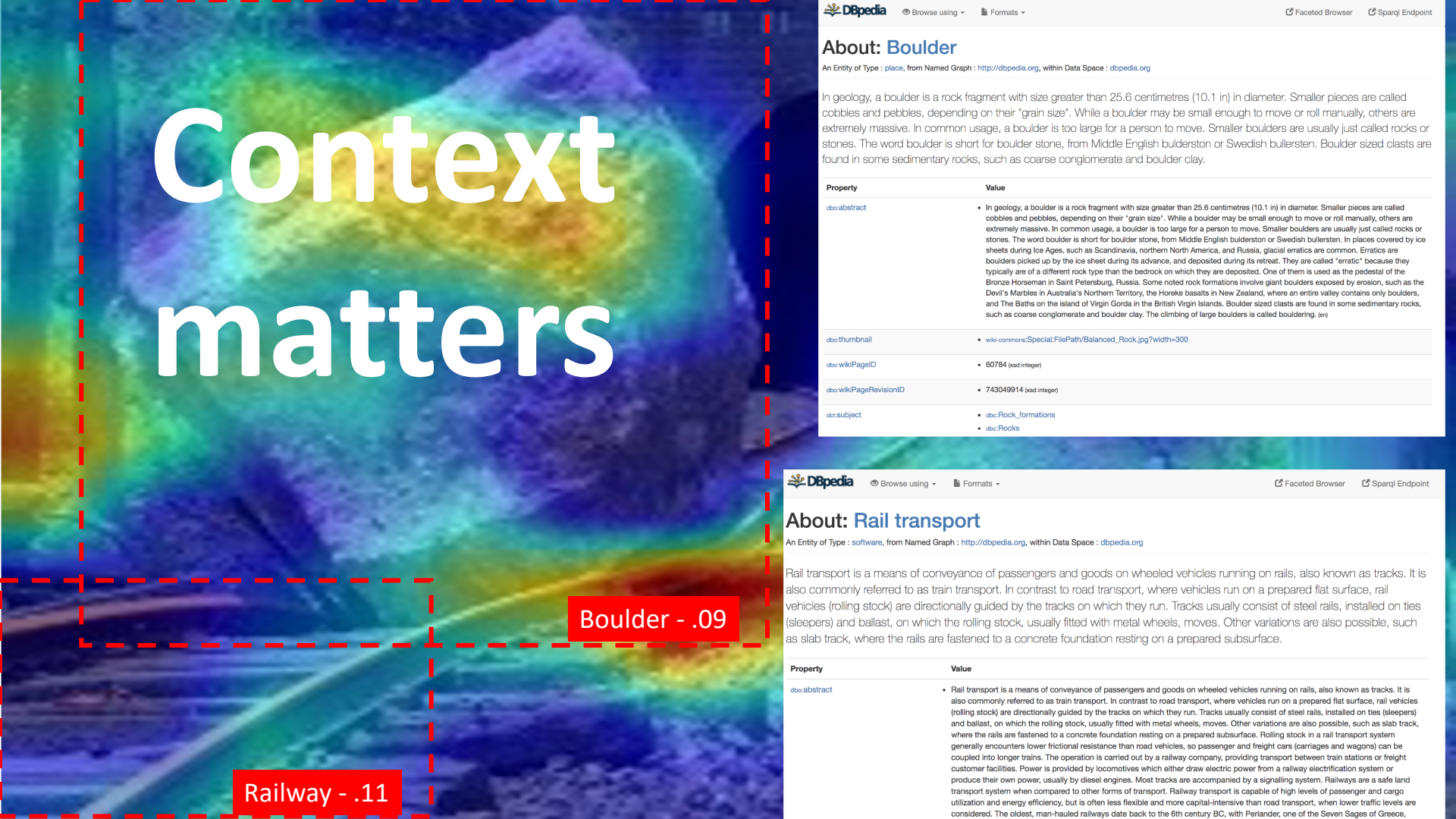
dbo:wikiPageID	352327 (xsd:integer)
dbo:wikiPageRevisionID	734430894 (xsd:integer)
dct:subject	dbc:Sawmills dbc:Saws dbc:Ancient_Roman_technology dbc:Timber_preparation dbc:Timber_industry
http://purl.org/linguistics/gold/hypernym	dbr:Facility
rdf:type	owl:Thing dbo:ArchitecturalStructure
rdfs:comment	A sawmill or lumber mill is a facility where logs are cut into lumber. Prior to the invention of the sawmill, boards were rived (split) and planed, or more often sawn by two men with a whipsaw, one above and another in a saw pit below. The earliest known mechanical mill is the Hierapolis sawmill, a Roman water-powered stone mill at Hierapolis, Asia Minor dating back to the 3rd century AD. Other water-powered mills followed and by the 11th century they were widespread in Spain and North Africa, the Middle East and Central Asia, and in the next few centuries, spread across Europe. The circular motion of the wheel was converted to a reciprocating motion at the saw blade. Generally, only the saw was powered, and the logs had to be loaded and moved by hand. An early improvement was the developm (en)
rdfs:label	Sawmill (en)
owl:sameAs	wikidata:Sawmill dbpedia-cs:Sawmill dbpedia-de:Sawmill dbpedia-es:Sawmill

OPEN

What is missing?

Lumbermill - .59





DBpedia Browse using Formats Faceted Browser Sparql Endpoint

About: Boulder

An Entity of Type : place, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

In geology, a boulder is a rock fragment with size greater than 25.6 centimetres (10.1 in) in diameter. Smaller pieces are called cobbles and pebbles, depending on their "grain size". While a boulder may be small enough to move or roll manually, others are extremely massive. In common usage, a boulder is too large for a person to move. Smaller boulders are usually just called rocks or stones. The word boulder is short for boulder stone, from Middle English bulderston or Swedish bullersten. Boulder sized clasts are found in some sedimentary rocks, such as coarse conglomerate and boulder clay.

Property	Value
dbo:abstract	▪ In geology, a boulder is a rock fragment with size greater than 25.6 centimetres (10.1 in) in diameter. Smaller pieces are called cobbles and pebbles, depending on their "grain size". While a boulder may be small enough to move or roll manually, others are extremely massive. In common usage, a boulder is too large for a person to move. Smaller boulders are usually just called rocks or stones. The word boulder is short for boulder stone, from Middle English bulderston or Swedish bullersten. In places covered by ice sheets during Ice Ages, such as Scandinavia, northern North America, and Russia, glacial erratics are common. Erratics are boulders picked up by the ice sheet during its advance, and deposited during its retreat. They are called "erratic" because they typically are of a different rock type than the bedrock on which they are deposited. One of them is used as the pedestal of the Bronze Horseman in Saint Petersburg, Russia. Some noted rock formations involve giant boulders exposed by erosion, such as the Devil's Marbles in Australia's Northern Territory, the Horeke basalts in New Zealand, where an entire valley contains only boulders, and The Baths on the island of Virgin Gorda in the British Virgin Islands. Boulder sized clasts are found in some sedimentary rocks, such as coarse conglomerate and boulder clay. The climbing of large boulders is called bouldering. (en)
dbo:thumbnail	▪ wiki-commons:Special:FilePath/Balanced_Rock.jpg?width=300
dbo:wikiPageID	▪ 60784 (xsd:integer)
dbo:wikiPageRevisionID	▪ 743049914 (xsd:integer)
dct:subject	▪ dbc:Rock_formations ▪ dbc:Rocks

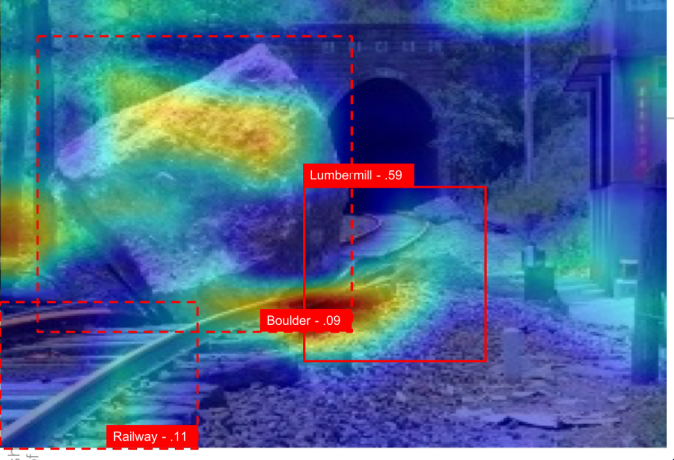
DBpedia Browse using Formats Faceted Browser Sparql Endpoint

About: Rail transport

An Entity of Type : software, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Rail transport is a means of conveyance of passengers and goods on wheeled vehicles running on rails, also known as tracks. It is also commonly referred to as train transport. In contrast to road transport, where vehicles run on a prepared flat surface, rail vehicles (rolling stock) are directionally guided by the tracks on which they run. Tracks usually consist of steel rails, installed on ties (sleepers) and ballast, on which the rolling stock, usually fitted with metal wheels, moves. Other variations are also possible, such as slab track, where the rails are fastened to a concrete foundation resting on a prepared subsurface.

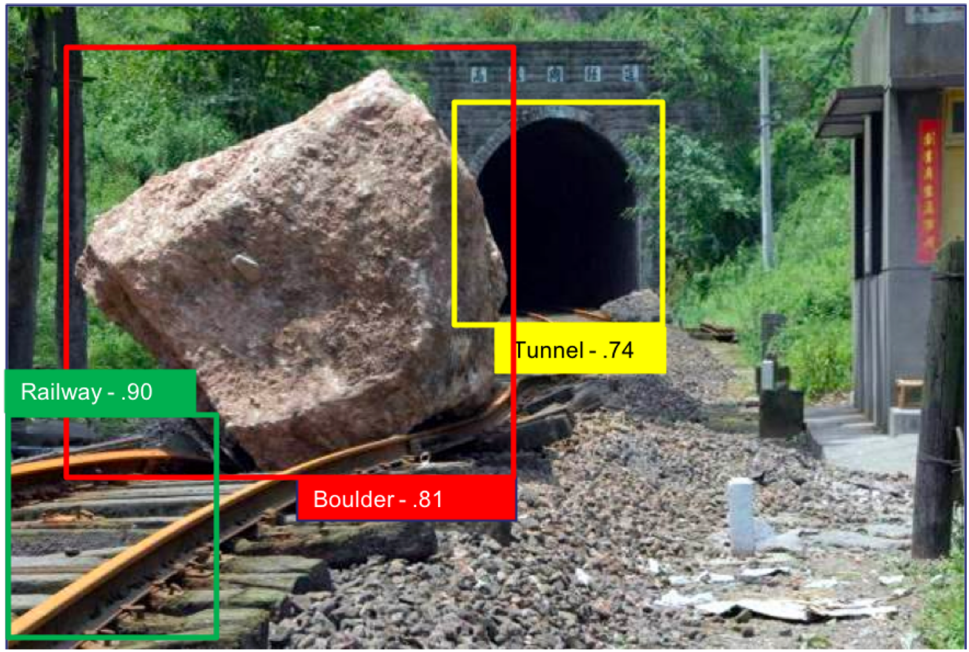
Property	Value
dbo:abstract	▪ Rail transport is a means of conveyance of passengers and goods on wheeled vehicles running on rails, also known as tracks. It is also commonly referred to as train transport. In contrast to road transport, where vehicles run on a prepared flat surface, rail vehicles (rolling stock) are directionally guided by the tracks on which they run. Tracks usually consist of steel rails, installed on ties (sleepers) and ballast, on which the rolling stock, usually fitted with metal wheels, moves. Other variations are also possible, such as slab track, where the rails are fastened to a concrete foundation resting on a prepared subsurface. Rolling stock in a rail transport system generally encounters lower frictional resistance than road vehicles, so passenger and freight cars (carriages and wagons) can be coupled into longer trains. The operation is carried out by a railway company, providing transport between train stations or freight customer facilities. Power is provided by locomotives which either draw electric power from a railway electrification system or produce their own power, usually by diesel engines. Most tracks are accompanied by a signalling system. Railways are a safe land transport system when compared to other forms of transport. Railway transport is capable of high levels of passenger and cargo utilization and energy efficiency, but is often less flexible and more capital-intensive than road transport, when lower traffic levels are considered. The oldest, man-hauled railways date back to the 6th century BC, with Periander, one of the Seven Sages of Greece,



- **Hardware**: **High performance, scalable, generic** (to different FGPA family) **& portable** CNN dedicated **programmable** processor implemented on an FPGA for **real-time embedded inference**

- **Software**: Knowledge graph extension of object detection

Transitioning



This is an **Obstacle: Boulder** obstructing the train: XG142-R on **Rail_Track** from City: Cannes to City: Marseille at **Location: Tunnel VIX** due to **Landslide**

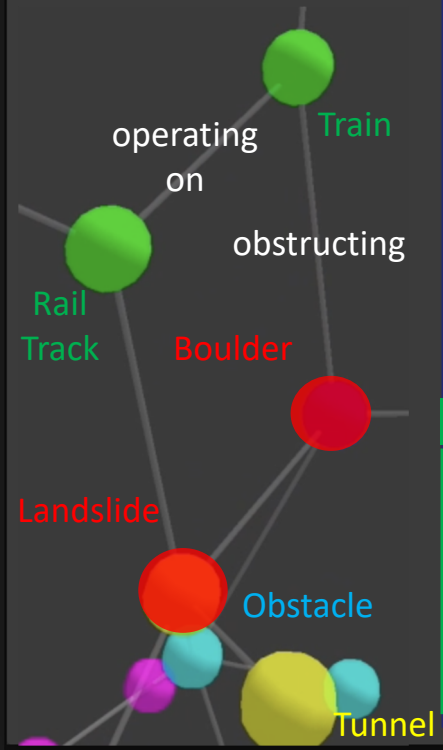
XAI Thales Platform

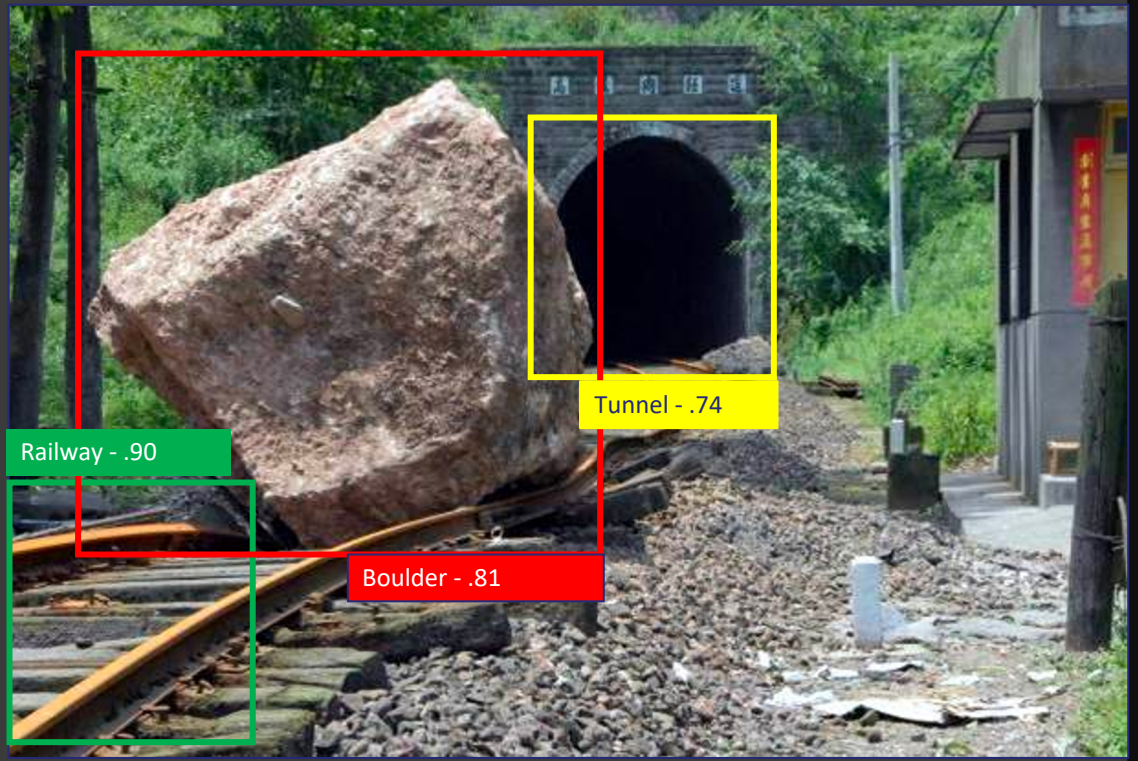
Higher accuracy with no intensive fine-tuning
Human interpretable explanation
Running on the edge at inference time

EXPLANATIONS

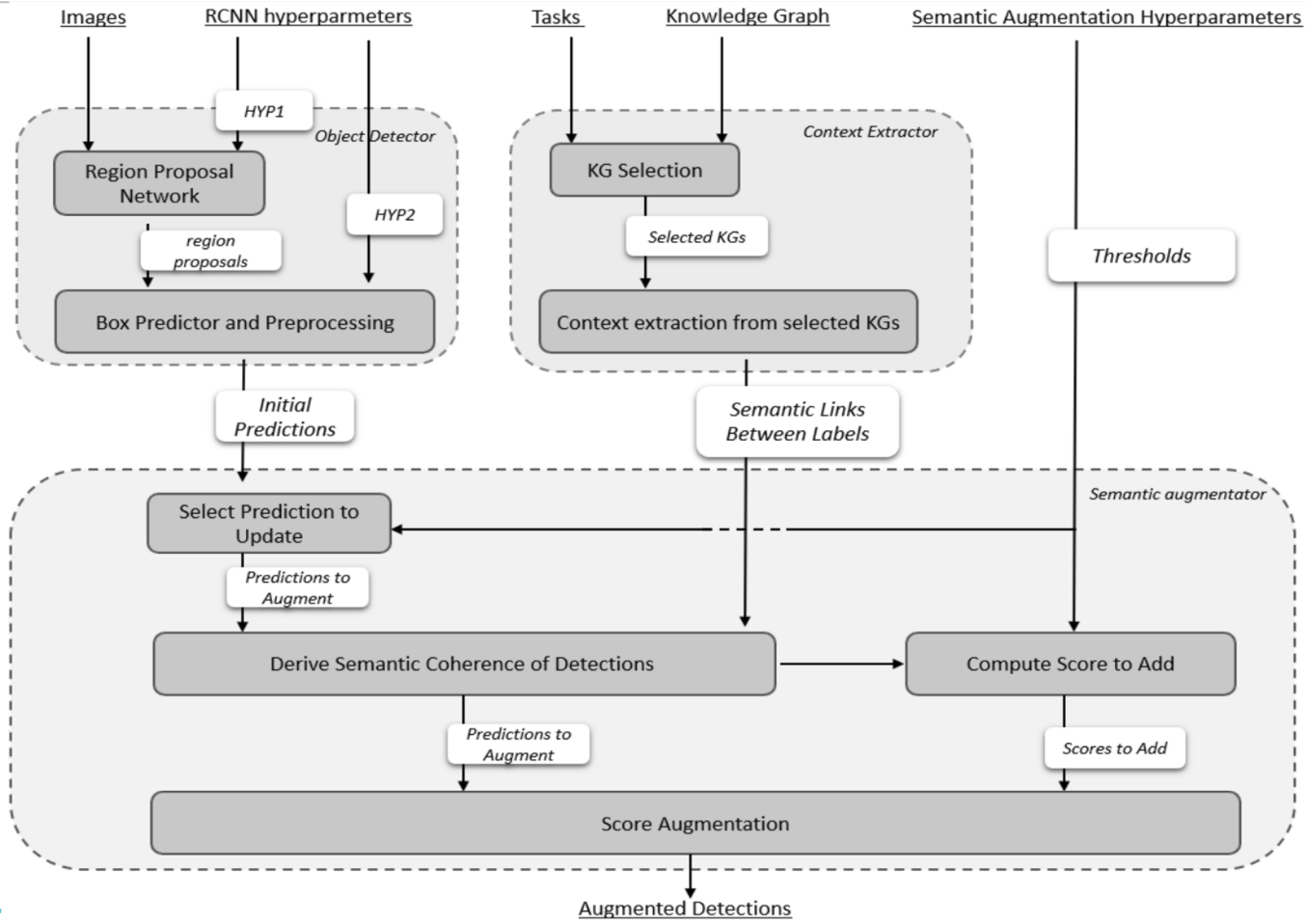
Lime

ResNet50 image classifier





Knowledge Graph in Machine Learning - An Implementation



Freddy Lécué, Jiaoyan Chen, Jeff Z. Pan, Huajun Chen: Augmenting Transfer Learning with Semantic Reasoning. IJCAI 2019: 1779-1785

Freddy Lécué, Tanguy Pommellet: Feeding Machine Learning with Knowledge Graphs for Explainable Object Detection. ISWC Satellites 2019: 277-280

Freddy Lécué, Baptiste Abeloos, Jonathan Anctil, Manuel Bergeron, Damien Dalla-Rosa, Simon Corbeil-Letourneau, Florian Martet, Tanguy Pommellet, Laura Salvan, Simon Veilleux, Maryam Ziaeefard: Thales XAI Platform: Adaptable Explanation of Machine Learning Systems - A Knowledge Graphs Perspective. ISWC Satellites 2019: 315-316

Jiaoyan Chen, Freddy Lécué, Jeff Z. Pan, Ian Horrocks, Huajun Chen: Knowledge-Based Transfer Learning Explanation. KR 2018: 349-358

THALES

Ecosystem

XAI in Canada: An instance of a Scientific eXplainable AI project

DEEL (Dependable Explainable Learning) Project 2019-2024

With support of

CRSNG NSERC

CRIAQ

Industrial partners

THALES

And 3 aerospace partners

Academic partners

UNIVERSITÉ LAVAL

IVADO

IRT SAINT EXUPERY

Certificability

- Structural warranties
- Risk auto evaluation
- External audit


ABB ROBOTICS
GoodAI
Allianz
BLUMORPHO
MŰEGYETEM 1782
LAAS-CNRS
CSIC
FRANCE DIGITALE
BSC
CARTIF
Atos
PGWC
Inria
eit Digital
brgm
Eötvös Loránd University
FORTH
INSTITUTE OF COMPUTER SCIENCE
cea
CINECA
DFKI
DLR
FORUM VIRIUM HELSINKI
GRASSROOTS ARTS
CERTH
CENTRE FOR RESEARCH & TECHNOLOGY HELLAS
INDUSTRIAL DATA SPACE ASSOCIATION
EOS
FundingBox
FONDAZIONE BRUNO KESSLER
Fraunhofer IAIS
UNIVERSITY OF LEEDS
IJS
Institut Mines-Télécom
TÉCNICO LISBOA
KIT
Karlsruhe Institute of Technology
KNOW Center
idiap
HUB FRANCE IA
orange
SAPIENZA UNIVERSITÀ DI ROMA
NTNU
Norwegian University of Science and Technology
Insight
Centre for Data Analytics
NUI Galway OF Gaillimh
ONERA
SmartRural
ÖREBRO UNIVERSITET
WAVESTONE
PANTHÉON SORBONNE
Qwant
SIEMENS
Ingenuity for life
SAP
SMILE
IT IS OPEN
simula
ThalesAlenia Space
technicolor
telenor
TILDE
THALES
Unilever
UMEÅ UNIVERSITET
TUM
ALMA MATER STUDIORUM
TU WIEN
UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh
UNIVERSIDADE DE COIMBRA
UNIVERSITÉ Grenoble Alpes
VUB
HELLENIC REPUBLIC
National and Kapodistrian University of Athens
UPC
POLITÉCNICA
UNIVERSITÀ DI SIENA
SCIENCES SORBONNE UNIVERSITÉ
eclt
European Centre for Living Technology

Human -in-the- Loop

Journées de Statistique de la SFdS, 2021, June 11th

Freddy Lecue, Chief AI Scientist (Thales Canada, Inria France)

@freddylecue