

Explainable Abstract Trains Dataset

Manuel de Sousa Ribeiro, Ludwig Krippahl, and Joao Leite

NOVA LINCS
School of Science and Technology
NOVA University Lisbon
Portugal

Abstract. The Explainable Abstract Trains Dataset is an image dataset containing simplified representations of trains. It aims to provide a platform for the application and research of algorithms for justification and explanation extraction. The dataset is accompanied by an ontology that conceptualizes and classifies the depicted trains based on their visual characteristics, allowing for a precise understanding of how each train was labeled. Each image in the dataset is annotated with multiple attributes describing the trains' features and with bounding boxes for the train elements.

1 Introduction

We introduce the Explainable Abstract Trains Dataset¹ (XTRAINS), an annotated image dataset focused on explainability, built with the objective of facilitating the research of algorithms for justification and explanation extraction. The XTRAINS dataset contains 500 000 images of 152×152 pixel art representations of trains, inspired by the trains developed by J. Larson and R. S. Michalski in [2], as shown in Figure 1.

The XTRAINS dataset was built for and used in [1] to illustrate a method to produce symbolic justifications for the output of artificial neural networks. It is accompanied by an ontology, shown in Figure 6 (in Appendix A), designed to conceptualize and describe the train representations in this dataset. The ontology provides ground-truth knowledge regarding how each image was labeled, allowing for ontology learning methods, such as the ones described in [4] to be benchmarked in this dataset. The provided bounding boxes allow for attribution methods [6,8] and salience mapping [7,9] methods to be tested, using evaluation metrics, e.g., like the one described in [5].

2 Image Composition

Each image in the dataset portrays a single train representation over a background image. All background images are random fragments from the images in the McGill Calibrated Colour Image Database [3]. A train is composed by

¹ Available at <https://bitbucket.org/xtrains/dataset/>



Fig. 1: Sample images of trains' representations.

a locomotive and a set of wagons that may have some contents. Each train representation is characterized as follows:

- **Wagons** The set of wagons contained in the train, further described below. The number of wagons in a train is equal to the integer part of a value sampled from a truncated normal distribution constrained to the interval $[1, 5]$, with a mean of 3, and a standard deviation of 1.
- **Wheel size** The size of each wheel in the train, one of 4 possible sizes. The probability mass function of a train's wheel size is shown in Figure 7 (in Appendix B).
- **Couplers' height** The height at which each coupler is attached, one of 2 possible heights. The height is uniformly distributed between those values.
- **Wagons' spacing** The number of pixels separating each wagon, one of 2 possible values. The separation is uniformly distributed between those values.
- **Position and angle** The trains' position and angle inside an image. A train's position is randomly selected taking into consideration the train's dimensions to ensure that the train is always visible in the image. The angle of each train is sampled from a truncated normal distribution constrained to the interval $[-30, 30]$, with a mean of 0, and a standard deviation of 9.

The effect of varying the wheel size, couplers' height, wagons' spacing and the position and angle of a train is shown, respectively, in each row of Figure 2.

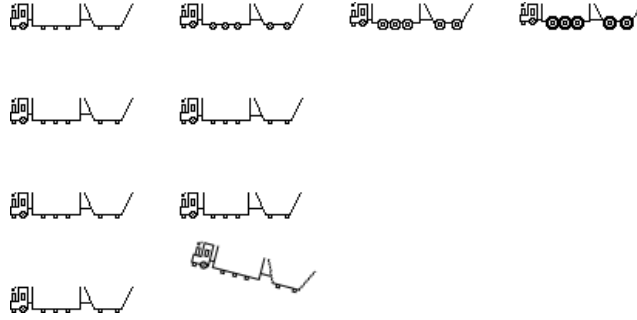


Fig. 2: Effect of individually varying trains' features.

A wagon is characterized as follows:

- **Length** The wagon's length in pixels. A wagon has a length of 34 pixels with a probability of $1/3$ and a length of 24 pixels with a probability of $2/3$.
- **Wall shape** The shape of the wagon's walls, one of 6 available shapes.
- **Roof shape** The shape of the wagon's roof, one of 5 available shapes.
- **Amount of visible wheels** The number of visible wheels in a wagon, one of 3 possible values.
- **Contents** The contents carried by a wagon, further described below.

The values of a wagons' wall shape, roof shape and amount of wheels are conditioned by the wagon's length and by each others' values, with probability mass functions of these features presented, respectively, in Figure 8, 9 and 10 (in Appendix B). The effect of individually varying each feature is depicted in Figure 3.

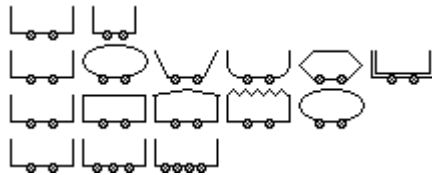


Fig. 3: Effect of individually varying wagons' features. First row varies length; second row varies wall shape; third row varies roof shape; fourth row varies the amount of visible wheels.

The contents of a wagon, if any, are characterized as follows:

- **Shape** The shape of the content, one of 8 available shapes.
- **Size** The size of the content, one of 4 available sizes.
- **Quantity** The content’s quantity, one of 3 possible values.
- **Position** The position of the contents inside a wagon. This value is dependent on the contents’ shape, size, and quantity and on the trains’ length and wheel size, ensuring that the contents are always inside the wagon.

The probabilities associated with the values of the shape, size, and quantity of a wagons’ content are conditioned by the wagon’s length, wall shape, and roof shape and by each others’ values, and can be seen, respectively, in Figures 11, 12, and 13 (in Appendix B). The effect of varying each of these features individually is illustrated, respectively, in each row of Figure 4.

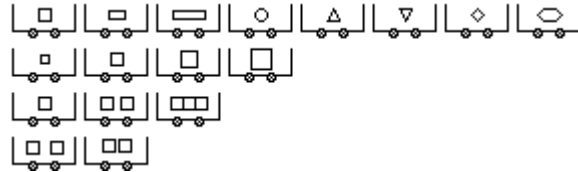


Fig. 4: Effect of individually varying wagons’ contents’ features. First row varies shape; second row varies size; third row varies quantity; fourth row varies position.

Additionally, we associate to the floor and each wall of a wagon a 15% probability of it being thicker, drawing it with a thickness of 2 pixels, instead of 1. Noise was also deliberately introduced in the form of missing pixels from the trains’ representations, up to 10% of the trains’ pixels might not be drawn. The resulting effect is depicted in Figure 5, the first train has no thick walls or missing pixels and the fourth train shows both effects combined.



Fig. 5: Addition of noise to the leftmost train representation, the second and third trains show, respectively, the effect of individually adding the possibility of thick walls and missing pixels to the first train.

3 Dataset Annotations

All images in the dataset are annotated with 39 binary attributes, shown in Figure 14 (in Appendix C), such as `TwoWheelsWagon` indicating if there is any wagon with 2 wheels visible in the image, and with 8 numeric attributes, shown

in Figure 15 (in Appendix C), like `Angle` indicating the angle of the train inside an image. The values for those attributes were either obtained directly when generating each image, e.g., the value of the attribute `NumberOfWagons`, indicating the number of wagons in an image, or obtained through reasoning with the axioms in the datasets’ ontology, e.g., `LongTrain`, indicating a train with, at least, 2 long wagons or 3 wagons. Furthermore, each image was also annotated with bounding boxes for the following elements:

- Train;
- Train’s Wagon;
- Wagon’s Contents;
- Wagon’s Roof;
- Wagon’s Wheels.

4 Image Generator

We also make available an image generator² capable of generating the XTRAINS dataset. This generator has multiple parameters, e.g., to define the length of the trains’ wagons, or to define the different possible amounts of wheels of a wagon, which are set to be in accord with the descriptions provided above. By changing the values of these parameters, it is possible to generate new datasets with different characteristics.

5 Conclusion

The XTRAINS dataset has a total 500 000 images over multiple different classes of trains. We hope that the large number of different annotations provided, coupled together with the knowledge embedded in the provided ontology make the XTRAINS dataset useful to benchmark different methods and techniques, notably in the area of Explainable AI.

Acknowledgements The authors would like to thank the support provided by Calouste Gulbenkian Foundation through its *Young Talents in AI* program, by FCT project ABSOLV (PTDC/CCI-COM/28986/2017), and by FCT strategic project NOVA LINCS (UIDB/04516/2020).

References

1. Manuel de Sousa Ribeiro and Joao Leite. Aligning artificial neural networks and ontologies towards explainable AI. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI’21*. AAAI Press, 2021.
2. J. Larson and Ryszard S. Michalski. Inductive inference of VL decision rules. *SIGART Newsl.*, 63:38–44, 1977.

² Available at <https://bitbucket.org/xtrains/dataset/>

3. Adriana Olmos and Frederick A. A. Kingdom. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, 33(12):1463–1473, 2004. PMID: 15729913.
4. Ana Ozaki. Learning description logic ontologies: Five approaches. where do they stand? *Künstliche Intell.*, 34(3):317–327, 2020.
5. Karl Schulz, Leon Sixt, Federico Tombari, and Tim Landgraf. Restricting the flow: Information bottlenecks for attribution. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.
6. Avanti Shrikumar, Peyton Greenside, Anna Shcherbina, and Anshul Kundaje. Not just a black box: Learning important features through propagating activation differences. *CoRR*, abs/1605.01713, 2016.
7. Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Workshop Track Proceedings*, 2014.
8. Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In David J. Fleet, Tomás Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I*, volume 8689 of *Lecture Notes in Computer Science*, pages 818–833. Springer, 2014.
9. Bolei Zhou, Aditya Khosla, Àgata Lapedriza, Aude Oliva, and Antonio Torralba. Object detectors emerge in deep scene cnns. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

A Ontology

$$\begin{aligned}
 & \text{WarTrain} \sqcup \text{EmptyTrain} \equiv \text{TypeA} \\
 & \text{PassengerTrain} \sqcup \text{LongFreightTrain} \equiv \text{TypeB} \\
 & \text{RuralTrain} \sqcup \text{MixedTrain} \equiv \text{TypeC} \\
 & \exists \text{has.}(\text{Wagon} \sqcup \text{Locomotive}) \equiv \text{Train} \\
 & \text{Train} \sqsubseteq (\leq 4 \text{ has.Wagon}) \\
 & \exists \text{has} \sqsubseteq \text{Train} \\
 & \exists \text{has}^{-1} \sqsubseteq \text{Wagon} \sqcup \text{Locomotive} \\
 & \text{Wagon} \sqcap \text{Locomotive} \sqsubseteq \perp \\
 & \exists \text{has.ReinforcedCar} \sqcap \exists \text{has.PassengerCar} \sqsubseteq \text{WarTrain} \\
 & \forall \text{has.}(\text{EmptyWagon} \sqcup \text{Locomotive}) \sqcap \exists \text{has.EmptyWagon} \equiv \text{EmptyTrain} \\
 & \exists \text{has.}(\text{PassengerCar} \sqcap \text{LongWagon}) \sqcup (\geq 2 \text{ has.PassengerCar}) \sqsubseteq \text{PassengerTrain} \\
 & \text{LongTrain} \sqcap \text{FreightTrain} \equiv \text{LongFreightTrain} \\
 & (\geq 2 \text{ has.LongWagon}) \sqcup (\geq 3 \text{ has.Wagon}) \sqsubseteq \text{LongTrain} \\
 & (\geq 2 \text{ has.FreightWagon}) \sqsubseteq \text{FreightTrain} \\
 & \exists \text{has.EmptyWagon} \sqcap \exists \text{has.}(\text{PassengerCar} \sqcup \text{FreightWagon}) \sqcap \neg \exists \text{has.LongWagon} \sqsubseteq \text{RuralTrain} \\
 & \exists \text{has.FreightWagon} \sqcap \exists \text{has.PassengerCar} \sqcap \exists \text{has.EmptyWagon} \sqsubseteq \text{MixedTrain} \\
 & \text{EmptyWagon} \sqcup \text{PassengerCar} \sqcup \text{FreightWagon} \equiv \text{Wagon} \\
 & \text{ReinforcedCar} \sqsubseteq \text{Wagon} \\
 & \text{PassengerCar} \sqsubseteq \text{Wagon} \\
 & \text{EmptyWagon} \sqsubseteq \text{Wagon} \\
 & \text{LongWagon} \sqsubseteq \text{Wagon} \\
 & \text{FreightWagon} \sqsubseteq \text{Wagon} \\
 & \text{OpenRoofCar} \sqsubseteq \text{Wagon} \\
 & \text{EmptyWagon} \sqcap \text{PassengerCar} \sqsubseteq \perp \\
 & \text{EmptyWagon} \sqcap \text{FreightWagon} \sqsubseteq \perp \\
 & \text{PassengerCar} \sqcap \text{FreightWagon} \sqsubseteq \perp
 \end{aligned}$$

Fig. 6: Ontology describing how the trains' representations are classified.

B Train Features' Probability Mass Function

P(WheelSize = s)			
s = XS	s = S	s = M	s = L
0.1	0.45	0.4	0.05

Fig. 7: Probability mass function of a train's wheel size.

P(WallShape = s WagonLength = l)						
s = Box	s = Ellipse	s = Trapezoid	s = Superellipse	s = Hexagon	s = Reinforced	l
0.2	0.1	0.2	0.2	0.1	0.2	S
0.1	0.2	0.1	0.2	0.2	0.2	L

Fig. 8: Probability mass function of a wagon's wall shape.

P(RoofShape = r WallShape = s)					s
r = Open	r = Flat	r = Triangular	r = Saw	r = Ellipse	
0.2	0.3	0.2	0.3	0.0	Box
0.0	0.0	0.0	0.0	1.0	Ellipse
0.4	0.2	0.2	0.2	0.0	Trapezoid
0.55	0.15	0.15	0.15	0.0	Superellipse
0.0	1.0	0.0	0.0	0.0	Hexagon
0.25	0.25	0.25	0.25	0.0	Reinforced

Fig. 9: Probability mass function of a wagon's roof shape.

P(NumberOfWheels = n WagonLength = l, WheelSize = s)			l	s
n = 2	n = 3	n = 4		
0.1	0.9	0.0	S	XS
0.9	0.1	0.0	S	S
1.0	0.0	0.0	S	M
1.0	0.0	0.0	S	L
0.0	0.8	0.2	L	XS
0.5	0.4	0.1	L	S
0.5	0.5	0.0	L	M
0.5	0.5	0.0	L	L

Fig. 10: Probability mass function of a wagon's number of visible wheels.

P(ContentShape = t WagonLength = l, WallShape = s, RoofShape = r)									l	s	r
t = None	t = S_Square	t = M_Square	t = L_Square	t = Circle	t = Triangle	t = Inv_Triangle	t = Diamond	t = Hexagon			
0.2	0.14	0.11	0.0	0.11	0.11	0.11	0.11	0.11	S	Box	Open
0.2	0.13	0.11	0.0	0.15	0.1	0.1	0.1	0.11	S	Box	Flat
0.13	0.13	0.1	0.0	0.19	0.12	0.12	0.11	0.1	S	Box	Triangular
0.14	0.13	0.1	0.0	0.19	0.11	0.11	0.1	0.12	S	Box	Saw
0.14	0.12	0.11	0.0	0.19	0.11	0.11	0.11	0.11	S	Ellipse	Ellipse
0.2	0.12	0.1	0.0	0.1	0.12	0.12	0.12	0.12	S	Trapezoid	Open
0.15	0.12	0.12	0.0	0.15	0.12	0.11	0.11	0.12	S	Trapezoid	Flat
0.15	0.11	0.11	0.0	0.16	0.12	0.12	0.12	0.11	S	Trapezoid	Triangular
0.14	0.11	0.11	0.0	0.15	0.12	0.12	0.12	0.13	S	Trapezoid	Saw
0.23	0.1	0.1	0.0	0.13	0.11	0.11	0.11	0.11	S	Superellipse	Open
0.18	0.1	0.1	0.0	0.16	0.12	0.11	0.11	0.12	S	Superellipse	Flat
0.17	0.1	0.1	0.0	0.18	0.12	0.11	0.11	0.11	S	Superellipse	Triangular
0.16	0.12	0.1	0.0	0.17	0.11	0.11	0.11	0.12	S	Superellipse	Saw
0.15	0.13	0.1	0.0	0.1	0.13	0.13	0.13	0.13	S	Hexagon	Flat
0.19	0.15	0.11	0.0	0.11	0.11	0.11	0.11	0.11	S	Reinforced	Open
0.15	0.15	0.12	0.0	0.14	0.11	0.11	0.11	0.11	S	Reinforced	Flat
0.13	0.14	0.11	0.0	0.19	0.11	0.11	0.11	0.1	S	Reinforced	Triangular
0.14	0.14	0.11	0.0	0.19	0.1	0.1	0.1	0.12	S	Reinforced	Saw
0.14	0.12	0.12	0.11	0.1	0.1	0.1	0.1	0.11	L	Box	Open
0.1	0.12	0.12	0.11	0.14	0.1	0.1	0.1	0.11	L	Box	Flat
0.1	0.11	0.11	0.1	0.16	0.11	0.11	0.11	0.09	L	Box	Triangular
0.1	0.11	0.11	0.1	0.16	0.1	0.1	0.1	0.12	L	Box	Saw
0.1	0.11	0.1	0.1	0.17	0.11	0.11	0.1	0.1	L	Ellipse	Ellipse
0.15	0.11	0.1	0.1	0.1	0.11	0.11	0.11	0.11	L	Trapezoid	Open
0.1	0.11	0.11	0.11	0.13	0.11	0.11	0.11	0.11	L	Trapezoid	Flat
0.13	0.1	0.1	0.1	0.14	0.11	0.11	0.11	0.1	L	Trapezoid	Triangular
0.11	0.1	0.1	0.1	0.14	0.11	0.11	0.11	0.12	L	Trapezoid	Saw
0.2	0.09	0.09	0.09	0.13	0.1	0.1	0.1	0.1	L	Superellipse	Open
0.14	0.09	0.09	0.09	0.15	0.11	0.11	0.11	0.11	L	Superellipse	Flat
0.14	0.09	0.09	0.09	0.16	0.11	0.11	0.11	0.1	L	Superellipse	Triangular
0.12	0.09	0.09	0.09	0.16	0.11	0.11	0.11	0.12	L	Superellipse	Saw
0.11	0.11	0.1	0.1	0.1	0.12	0.12	0.12	0.12	L	Hexagon	Flat
0.14	0.12	0.12	0.11	0.1	0.1	0.1	0.1	0.11	L	Reinforced	Open
0.1	0.12	0.12	0.11	0.14	0.1	0.1	0.1	0.11	L	Reinforced	Flat
0.1	0.11	0.11	0.1	0.16	0.11	0.11	0.11	0.09	L	Reinforced	Triangular
0.1	0.11	0.11	0.1	0.16	0.1	0.1	0.1	0.12	L	Reinforced	Saw

Fig. 11: Probability mass function of a wagon's content shape.

P(ContentSize = s WagonLength = l, ContentShape = t, ContentQuantity = n)				l	t	n
XS	S	M	L			
0.4	0.3	0.3	0.0	S	S_Square, Circle, Triangle, Inv_Triangle, Diamond	1, 2, 3
1.0	0.0	0.0	0.0	S	M_Square	1, 2, 3
0.8	0.2	0.0	0.0	S	Hexagon	1, 2, 3
0.27	0.3	0.38	0.05	L	S_Square, Circle, Triangle, Inv_Triangle, Diamond	1, 2, 3
1.0	0.0	0.0	0.0	L	M_Square, L_Square	1, 2
0.3	0.4	0.3	0.0	L	Hexagon	1
1.0	0.0	0.0	0.0	L	Hexagon	2

Fig. 12: Probability mass function of a wagon's content size.

P(ContentQuantity = n WagonLength = l, ContentShape = t, ContentSize = s)			l	t	s
n = 1	n = 2	n = 3			
0.8	0.2	0.0	S	S_Square, Circle, Triangle, Inv_Triangle, Diamond	XS
1.0	0.0	0.0	S	S_Square, Circle, Triangle, Inv_Triangle, Diamond	S, M, L
1.0	0.0	0.0	S	M_Square, Hexagon	XS, S, M, L
0.3	0.3	0.4	L	S_Square, Circle, Triangle, Inv_Triangle, Diamond	XS
0.55	0.35	0.1	L	S_Square, Circle, Triangle, Inv_Triangle, Diamond	S
0.6	0.4	0.0	L	S_Square, Circle, Triangle, Inv_Triangle, Diamond	M
0.65	0.35	0.0	L	M_Square, Hexagon	XS, S, M, L
1.0	0.0	0.0	L	L_Square	XS, S, M, L

Fig. 13: Probability mass function of a wagon's content quantity.

C Dataset Attributes

TypeA	TypeB	TypeC	EmptyTrain
FreightTrain	LongFreightTrain	LongTrain	MixedTrain
PassengerTrain	RuralTrain	WarTrain	LongWagon
BoxCar	EllipseCar	HexagonCar	ReinforcedCar
SupperellipseCar	TrapezoidCar	FlatRoofCar	OpenRoofCar
SawRoofCar	TriangularRoofCar	TwoWheelsWagon	ThreeWheelsWagon
FourWheelsWagon	DiamondCargo	EmptyWagon	FreightWagon
HexagonCargo	InvTriangleCargo	PassengerCar	TriangleCargo
SSquareCargo	MSquareCargo	XLSquareCargo	XSWheelsTrain
SWheelsTrain	MWheelsTrain	LWheelsTrain	

Fig. 14: XTRAINS dataset's binary attributes.

Angle	CouplerHeight
NumberOfFreightWagons	NumberOfLongWagons
NumberOfPassengerCars	NumberOfWagons
TotalCargoQuantity	WagonSpacing

Fig. 15: XTRAINS dataset's numeric attributes.