



UNIVERSITEIT VAN AMSTERDAM



Representing and Explaining Novel Concepts with Minimal Supervision

Dr. Zeynep Akata

2 April 2019

Outline

Motivating the Importance of Side Information

(Generalized) Zero-Shot Learning with Side Information

Deeply Explainable Artificial Intelligence

Summary and Future Work

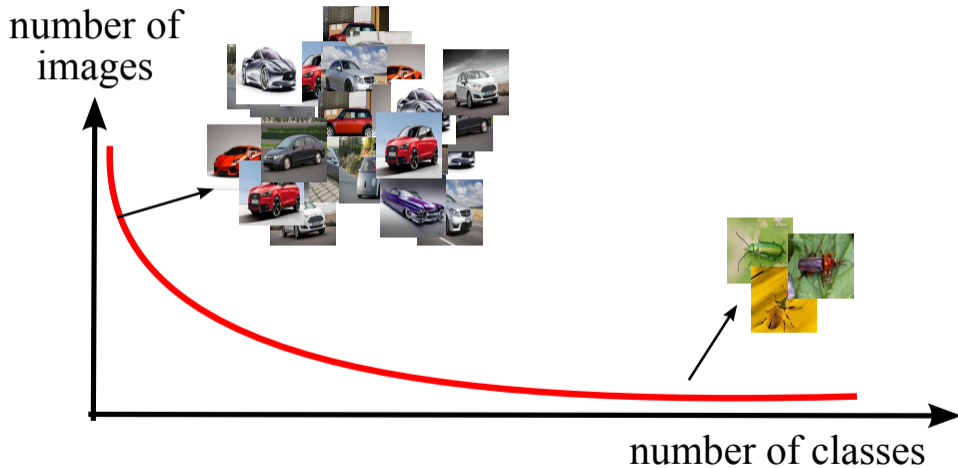
Outline

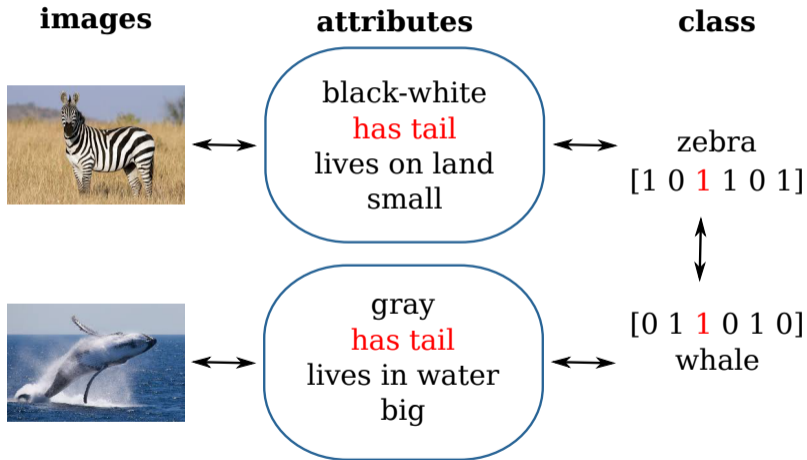
Motivating the Importance of Side Information

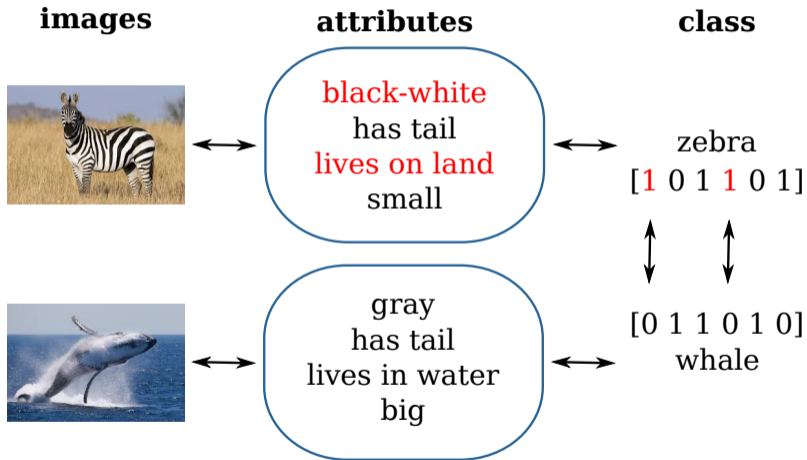
(Generalized) Zero-Shot Learning with Side Information

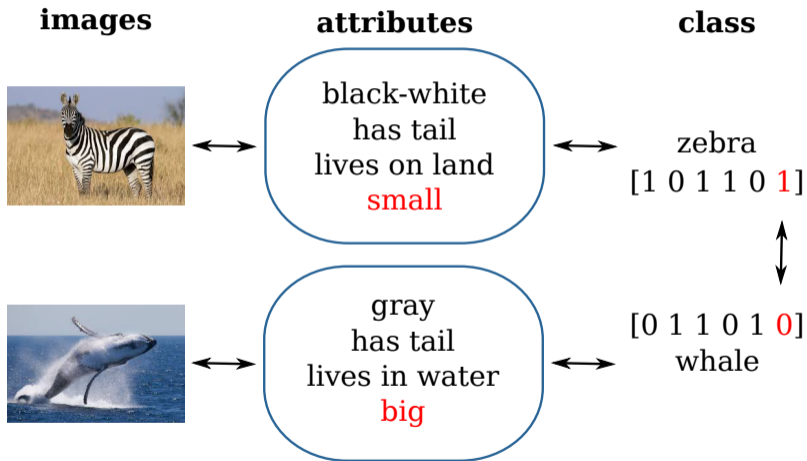
Deeply Explainable Artificial Intelligence

Summary and Future Work



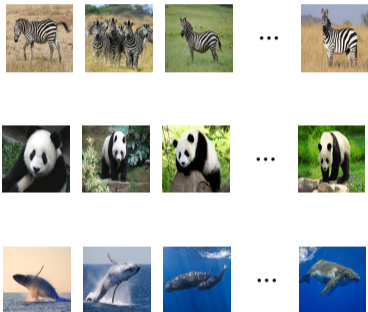






Zero-Shot Learning

images



attributes

black-white
has tail
lives on land
small

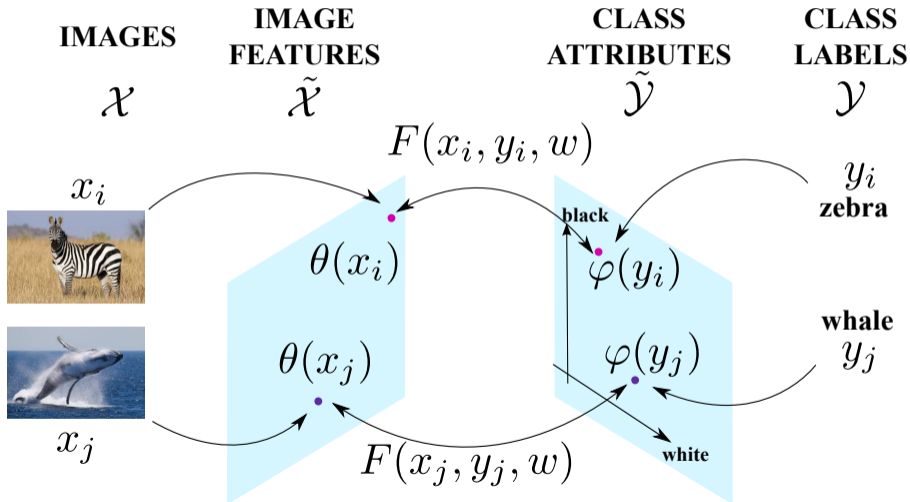
black-white
no tail
lives on land
medium

gray
has tail
lives in water
big

white
has tail
lives on land
tiny

Multimodal Embeddings

Akata et al. CVPR'13 & TPAMI'16



Zero-Shot Learning Datasets

Animals with
Attributes (AWA)

[Lampert et.al. CVPR'09]

50	85
cls	att



Caltech UCSD-Birds
(CUB)

[Wah et.al.'11]

200	312
cls	att



Zero-Shot vs Generalized Zero-Shot Learning

Xian et al. CVPR 2017

Method	Zero-Shot Learning		Generalized Zero-Shot Learning					
	CUB	AWA	CUB			AWA		
	u	u	u	s	H	u	s	H
Supervised Learning	-	-	-	82.1	-	-	96.2	-

Method	Zero-Shot Learning		Generalized Zero-Shot Learning					
	CUB	AWA	CUB			AWA		
	u	u	u	s	H	u	s	H
Supervised Learning	–	–	–	82.1	–	–	96.2	–
Multimodal Embeddings	54.9	59.9	23.7	62.8	34.4	16.8	76.1	27.5

Conclusions

Standard image classification models fail with the lack of labels

1. Zero-Shot Learning is a challenging task that deserves attention
2. Side information, e.g. attributes, is required to tackle zero-shot learning
3. Several sources of side information exist: moving towards free-form text

Akata et.al. IEEE CVPR 2013, 2015, 2016 & IEEE TPAMI 2014, 2016

Outline

Motivating the Importance of Side Information

(Generalized) Zero-Shot Learning with Side Information

Deeply Explainable Artificial Intelligence

Summary and Future Work

How to Tackle the Missing Data Problem?

Labels are difficult to obtain, attributes require expert knowledge

How to Tackle the Missing Data Problem?

Labels are difficult to obtain, attributes require expert knowledge

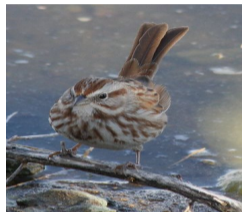
Proposed solution: Free text to image synthesis!

Detailed Visual Descriptions

Reed et al. CVPR'16



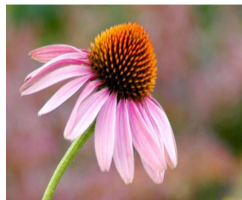
The bird has a white underbelly, black feathers in the wings, a large wingspan, and a white beak.



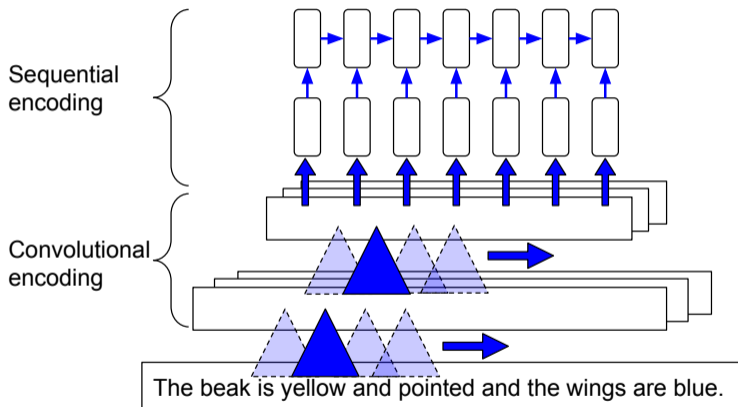
This bird has distinctive-looking brown and white stripes all over its body, and its brown tail sticks up.



This flower has a central white blossom surrounded by large pointed red petals which are veined and leaflike.



Light purple petals with orange and black middle green leaves



Text to Image Synthesis

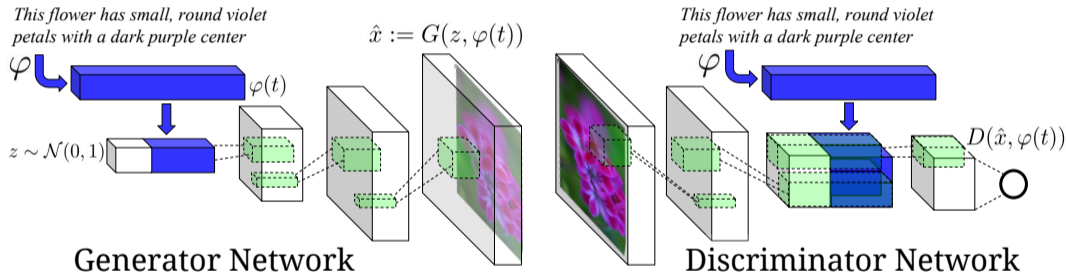
This large bird has **black feet** and
dark-brown feathers.



??

GAN¹ Conditioned on Text

Reed et al. ICML'16 & NIPS'16



¹Generative Adversarial Networks [Goodfellow et al. NIPS'14]

Text to Image Synthesis Results

Query a small sized bird that has tones of brown and dark red with a short stout bill

Generated Image



Query the bird has a yellow bill, pink webbed feet, a white body with gray wings and gray tail feathers

Generated Image



Query this bird is all blue, the top part of the bill is blue, but the bottom half is white.

Generated Image



Interpolating Between Sentences

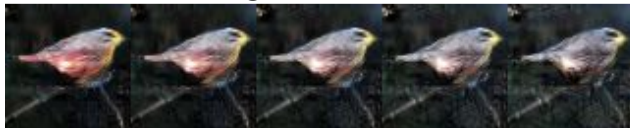
‘Blue bird with black beak’ →
‘Red bird with black beak’



‘Small blue bird with black wings’ →
‘Small yellow bird with black wings’



‘This bird is bright.’ → ‘This bird is dark.’



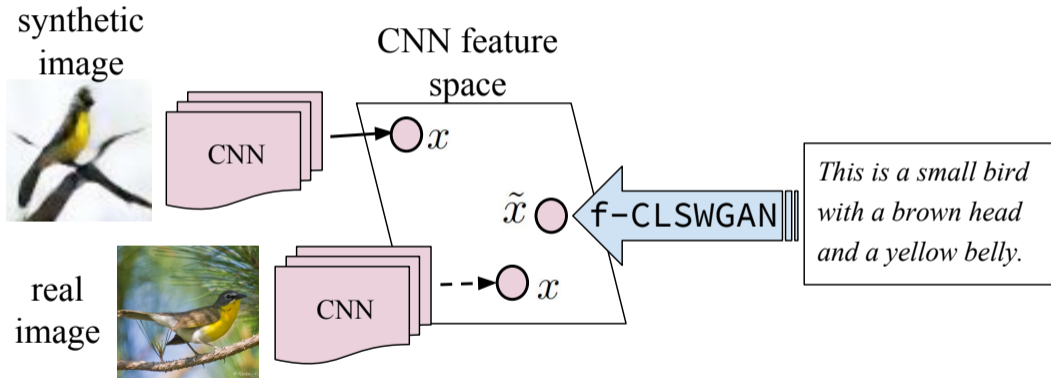
Generalized Zero-Shot Learning with Synthesized Images

Data	CUB		
	u	s	H
Only real data	23.7	62.8	34.4

Generalized Zero-Shot Learning with Synthesized Images

Data	CUB		
	u	s	H
Only real data	23.7	62.8	34.4
With generated images	23.8	48.5	31.9

This is not better than having no images!

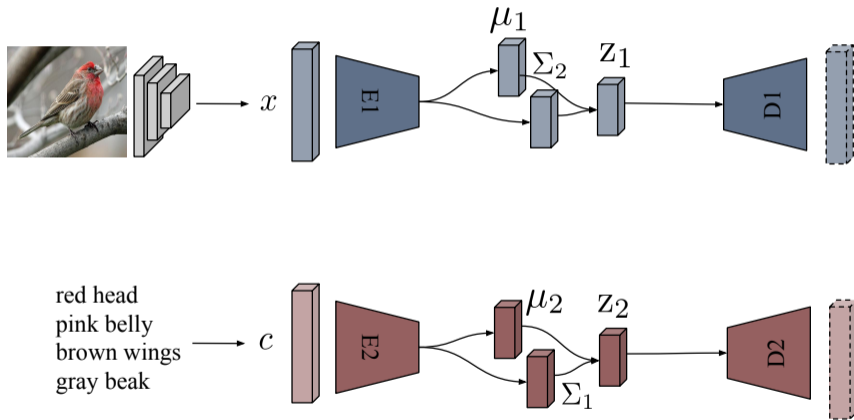


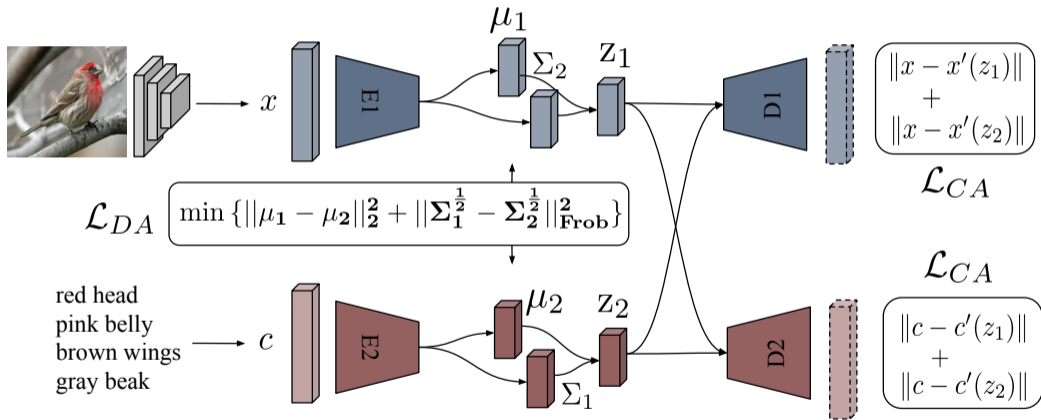
Generalized Zero-Shot Learning with Synthesized Image Features

Data	CUB		
	u	s	H
Only real data	23.7	62.8	34.4
With generated images	23.8	48.5	31.9

Generalized Zero-Shot Learning with Synthesized Image Features

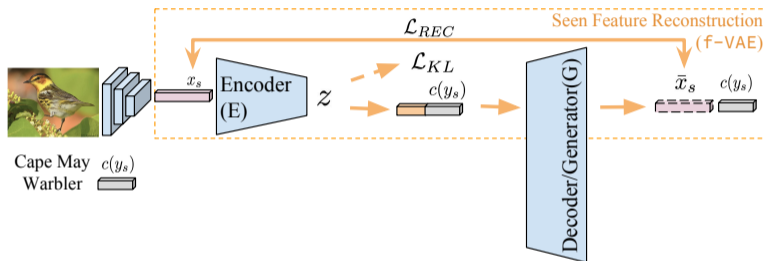
Data	CUB		
	u	s	H
Only real data	23.7	62.8	34.4
With generated images	23.8	48.5	31.9
With generated features (f-CLSWGAN)	43.7	57.7	49.7

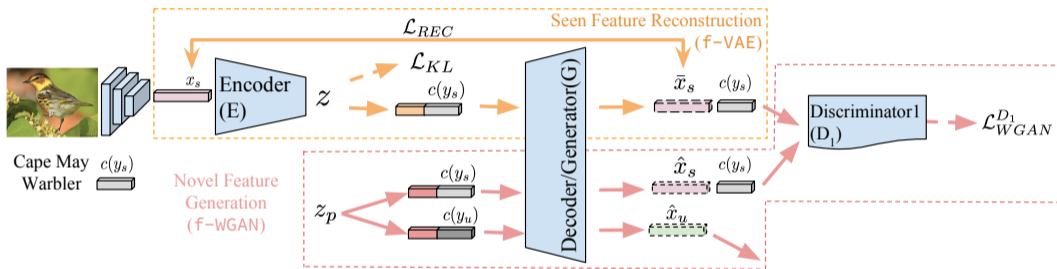




Generalized Zero-Shot Learning with Synthesized Image Features

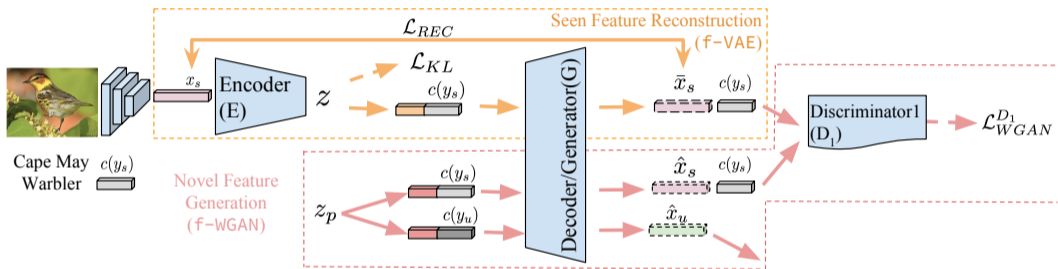
Data	CUB		
	u	s	H
Only real data	23.7	62.8	34.4
With generated images	23.8	48.5	31.9
With generated features (f-CLSWGAN)	43.7	57.7	49.7
With generated features (CADA-VAE)	63.6	51.6	52.4

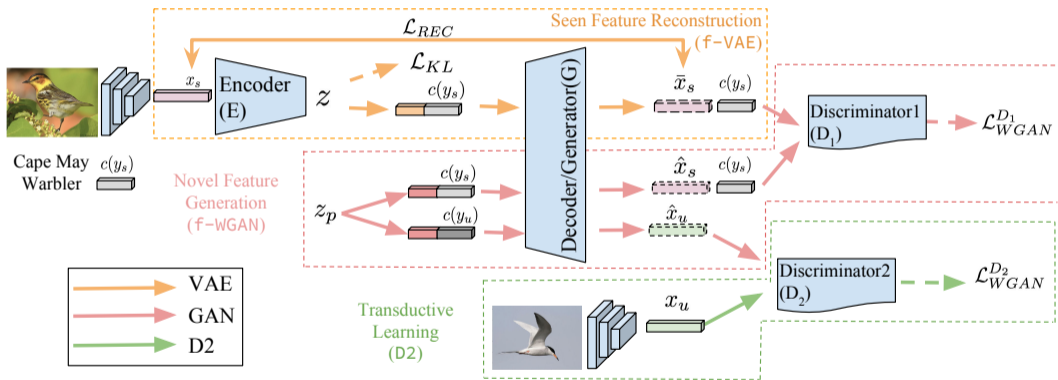




Generalized Zero-Shot Learning with Synthesized Image Features

Data	CUB		
	u	s	H
Only real data	23.7	62.8	34.4
With generated images	23.8	48.5	31.9
With generated features (f-CLSWGAN)	43.7	57.7	49.7
With generated features (CADA-VAE)	63.6	51.6	52.4
With generated features (f-VAEGAN-D2)	63.2	75.6	68.9





Generalized Zero-Shot Learning with Synthesized Image Features

Data	CUB		
	u	s	H
Only real data	23.7	62.8	34.4
With generated images	23.8	48.5	31.9
With generated features (f-CLSWGAN)	43.7	57.7	49.7
With generated features (CADA-VAE)	63.6	51.6	52.4
With generated features (f-VAEGAN-D2)	63.2	75.6	68.9
With generated features (f-VAEGAN-D2 tran)	73.8	81.4	77.3

Conclusions

Language complements visual information

1. Provides an intuitive interface for the model
2. Strong and generalizable: image classification & generation
3. Many potential applications: Towards explaining what deep models think

Akata et al. IEEE CVPR 2013, 2015, 2016, TPAMI 2014, 2016

Reed et al. IEEE CVPR 2016 & ICML 2016 & NIPS 2016

Xian et al. IEEE CVPR 2016, 2017, 2018, 2019

Schönfeld et al. IEEE CVPR 2019; Dutta and Akata IEEE CVPR 2019

Outline

Motivating the Importance of Side Information

(Generalized) Zero-Shot Learning with Side Information

Deeply Explainable Artificial Intelligence

Summary and Future Work

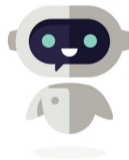
Human Machine Communication: Visual Question Answering



Human Machine Communication: Visual Question Answering



What type of bird is
this?



Human Machine Communication: Visual Question Answering



What type of bird is this?



It is a **Cardinal**



Human Machine Communication: Visual Question Answering



What type of bird is this?

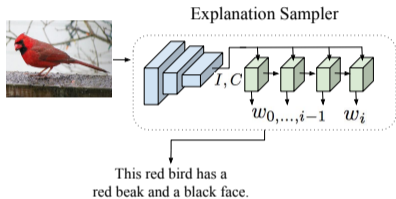


It is a **Cardinal** because it is a red bird with a red beak and a black face



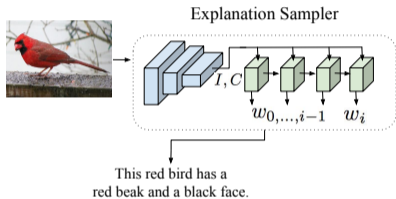
Generating Visual Explanations Model

Hendricks et al. ECCV'16



Generating Visual Explanations Model

Hendricks et al. ECCV'16



*This is a **Downy Woodpecker** because...*



D: this bird has a white breast black wings and a **red spot** on its head.

E: this is a black and white bird with a **red spot** on its crown.

*This is a **Downy Woodpecker** because...*

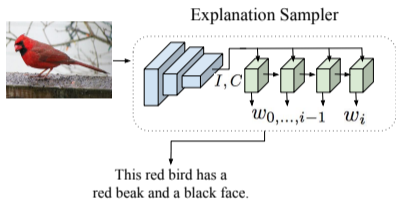


D: this bird has a white breast black wings and a **red spot** on its head.

E: this is a white bird with a black wing and a black and white striped head.

Generating Visual Explanations Model

Hendricks et al. ECCV'16



This is a Downy Woodpecker because...



D: this bird has a white breast black wings and a **red spot** on its head.
E: this is a black and white bird with a **red spot** on its crown.

This is a Downy Woodpecker because...



D: this bird has a white breast black wings and a **red spot** on its head.
E: this is a white bird with a black wing and a black and white striped head.

Correct: Laysan Albatross, **Predicted:** Cactus Wren



Explanation: ...this is a brown and white spotted bird with a long pointed beak.

Correct & Predicted: Laysan Albatross



Explanation: ...this bird has a white head and breast with a long hooked bill.

Cactus Wren Definition: ...this bird has a long thin beak with a brown body and black spotted feathers.

Laysan Albatross Definition: ...this bird has a white head and breast a grey back and wing feathers and an orange beak.

Human Machine Communication: Grounding & Counterfactuals



What type of bird is this?



It is a **Cardinal** because it is a red bird with a red beak and a black face



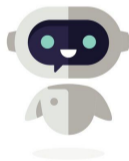
Human Machine Communication: Grounding & Counterfactuals



What type of bird is this?



It is a **Cardinal** because it is a **red bird** with a **red beak** and a **black face**



Human Machine Communication: Grounding & Counterfactuals

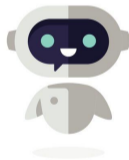


What type of bird is this?



Why not a **Vermilion Flycatcher**?

It is a **Cardinal** because it is a **red bird** with a **red beak** and a **black face**



Human Machine Communication: Grounding & Counterfactuals



What type of bird is this?



Why not a **Vermilion Flycatcher**?

It is a **Cardinal** because it is a **red bird** with a **red beak** and a **black face**

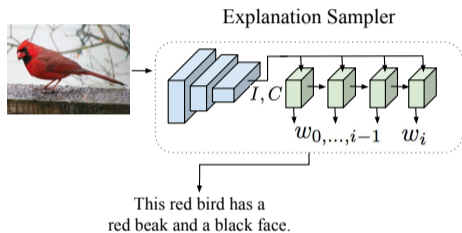


It is not a **Vermilion Flycatcher** because it does not have black wings.



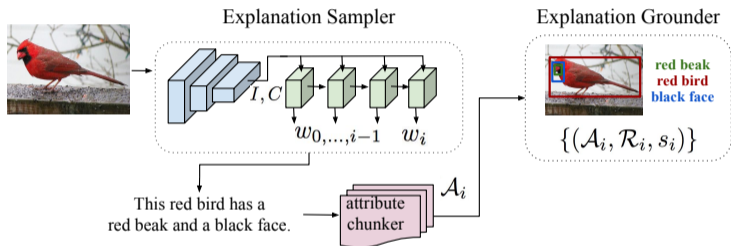
Grounding Visual Explanations

Hendricks et al. ECCV'18

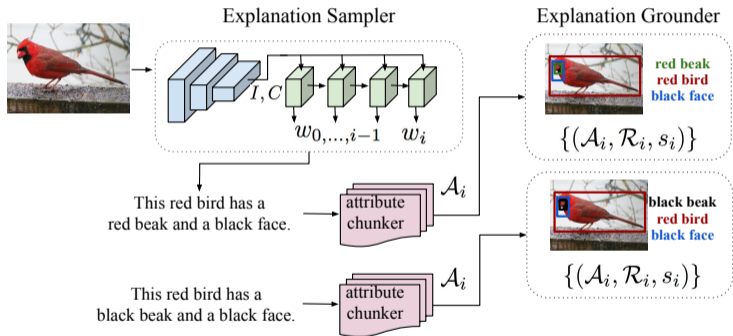


Grounding Visual Explanations

Hendricks et al. ECCV'18

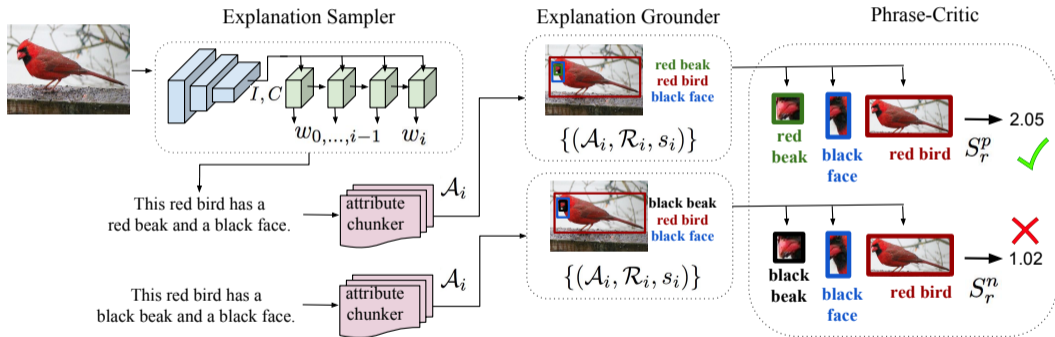


Grounding Visual Explanations



Grounding Visual Explanations

Hendricks et al. ECCV'18



Grounding Visual Explanations and Counterfactuals

This is a **Red Winged Blackbird** because



this is a **black bird** with a **red spot on its wingbars.**

Score: -11.29



this is a black bird with a red wing and a pointy black beak.

Grounding Visual Explanations and Counterfactuals

This is a **Red Winged Blackbird** because



this is a **black bird** with a **red spot on its wingbars**.

Score: -11.29



this is a black bird with a red wing and a pointy black beak.

Counterfactuals: Contrasting explanations are intuitive and informative



This bird is a **Crested Auklet** because this is a black bird with a small orange beak and it is not a **Red Faced Cormorant** because it does not have a long flat bill.

Textual Explanations for Self Driving Vehicles



What is
happening
here?



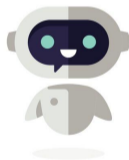
Textual Explanations for Self Driving Vehicles



What is happening here?



The car is moving forward.



Textual Explanations for Self Driving Vehicles



What is happening here?



Why do you think so?

The car is moving forward.



Textual Explanations for Self Driving Vehicles



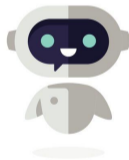
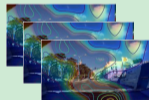
What is happening here?



Why do you think so?

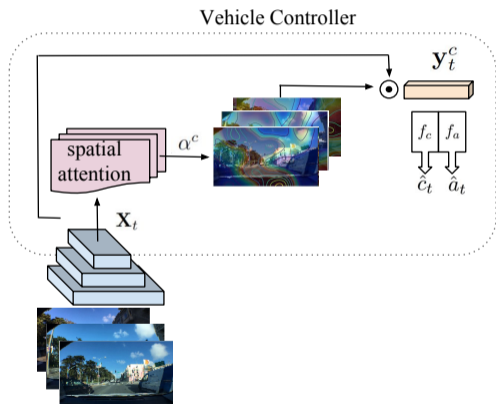
The car is moving forward.

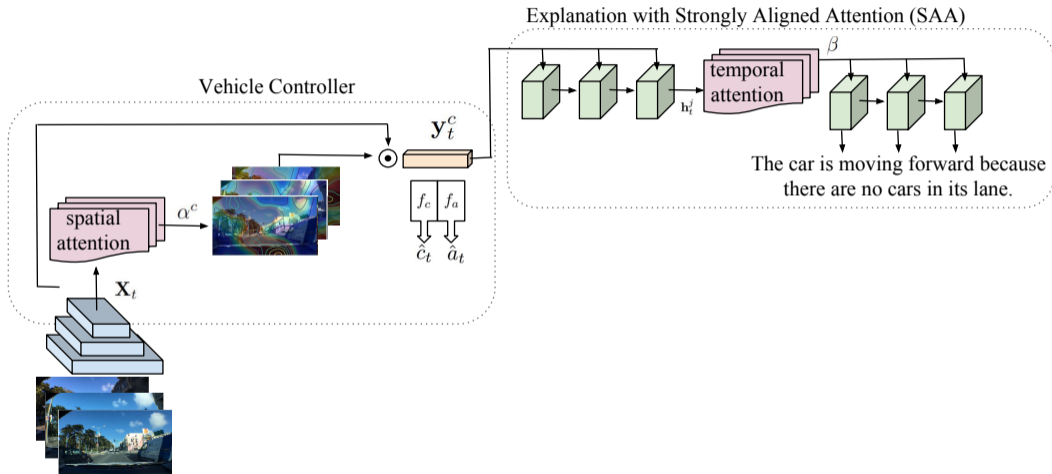
Because there are no cars in its lane.

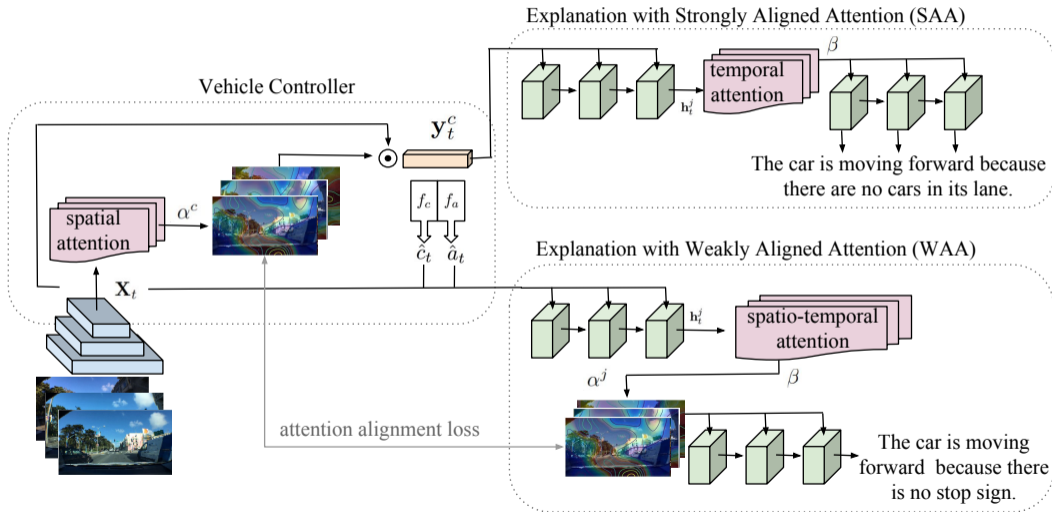


Driving Explanations Model

Kim et al. ECCV'18



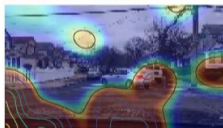




Driving Explanations Results



The car heads down the road because traffic is moving at a steady pace.



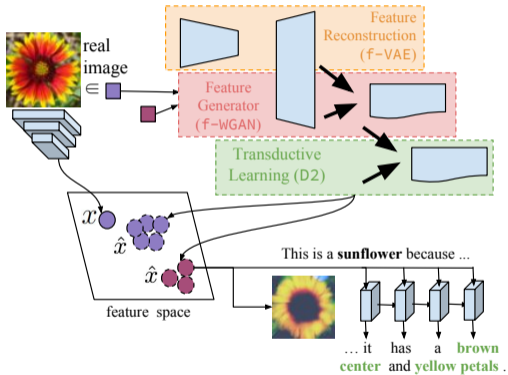
The car is slowing because it is approaching a stop sign.



The car is stopped because the car in front of it is stopped.

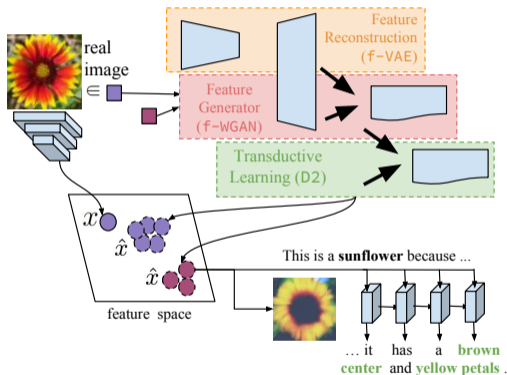
Zero-Shot Explanations

Xian et al. CVPR'19



Zero-Shot Explanations

Xian et al. CVPR'19



This is a **Tree Poppy** because ...

... the flower is **big with white petals**, and a **bulb of yellow colored anthers**.



... this flower has **petals that are white** and has a **bushy yellow center**



This is a **Purple Coneflower** because ...

... this flower has **pink petals that are pointed down**, and a **lot of red stamen in the center**



... this flower is **pink in color**, and has petals that are **drooping downward**.



This is a **Blanket Flower** because ...

... this flower has **petals that are red with yellow edges**



... this flower has **red petals** that have **yellow tips**.



This is a **Pink Primrose** because ...

... this flower is **pink and white in color**, with **petals that are rounded**.



... the **petals of the flower are light pink**, while the **anthers are white and yellow**.



Conclusions

Generating visual/textual explanations

1. A means for model interpretation: necessary to improve deep models
2. Important criteria to trust deep models: through explanations
3. A step towards effective human-machine communication

Hendricks et al. ECCV 2016 & ECCV 2018,
Park et al. IEEE CVPR 2018, Kim et al. ECCV 2018
Xian et al. IEEE CVPR 2019

Outline

Motivating the Importance of Side Information

(Generalized) Zero-Shot Learning with Side Information

Deeply Explainable Artificial Intelligence

Summary and Future Work

Summary

1. Multi-modal Joint Embeddings tackle lack of visual data
[Akata et al. TPAMI'14 & CVPR'13, CVPR'15 & TPAMI'16, Xian et al. CVPR'16]

Summary

1. Multi-modal Joint Embeddings tackle lack of visual data
[Akata et al. TPAMI'14 & CVPR'13, CVPR'15 & TPAMI'16, Xian et al. CVPR'16]
2. Vision and Language complement each other for generating novel concepts
[Reed et al. CVPR'16 & ICML'16 & NIPS'16, Xian et al. CVPR'18 & CVPR'19, Schönfeld et al. CVPR'19, Dutta and Akata CVPR'19]

Summary

1. Multi-modal Joint Embeddings tackle lack of visual data
[Akata et al. TPAMI'14 & CVPR'13, CVPR'15 & TPAMI'16, Xian et al. CVPR'16]
2. Vision and Language complement each other for generating novel concepts
[Reed et al. CVPR'16 & ICML'16 & NIPS'16, Xian et al. CVPR'18 & CVPR'19, Schönfeld et al. CVPR'19, Dutta and Akata CVPR'19]
3. Developing explainable deep models is important for user acceptance
[Hendricks et al. ECCV'16 & ECCV'18, Park et al. CVPR'18, Kim et al. ECCV'18, Xian et.al. CVPR'19]

Future of Deeply Explainable Artificial Intelligence



Future of Deeply Explainable Artificial Intelligence



User: What happened?

Future of Deeply Explainable Artificial Intelligence



User: What happened?

XAI: I was driving down an empty road. I decided to slow down as a ball appeared on the right. I saw a child running towards the ball, so I decided to stop.

Future of Deeply Explainable Artificial Intelligence



User: What happened?

XAI: I was driving down an empty road. I decided to slow down as a ball appeared on the right. I saw a child running towards the ball, so I decided to stop.

User: What would have happened if you did not stop ?

Future of Deeply Explainable Artificial Intelligence



User: What happened?

XAI: I was driving down an empty road. I decided to slow down as a ball appeared on the right. I saw a child running towards the ball, so I decided to stop.

User: What would have happened if you did not stop ?

XAI: If there was an impact, the child would have gotten hurt.

Thank you!