Histopathology and Deep Learning

Bryan Cardenas
Matthieu

SURF

Plan For Today

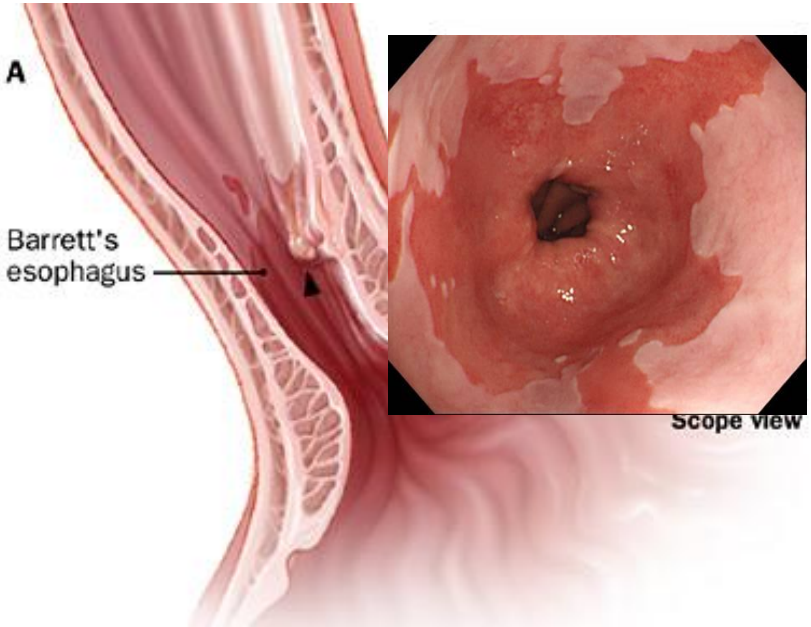
01. Histopathology Introduction

02. Examode: CLIP

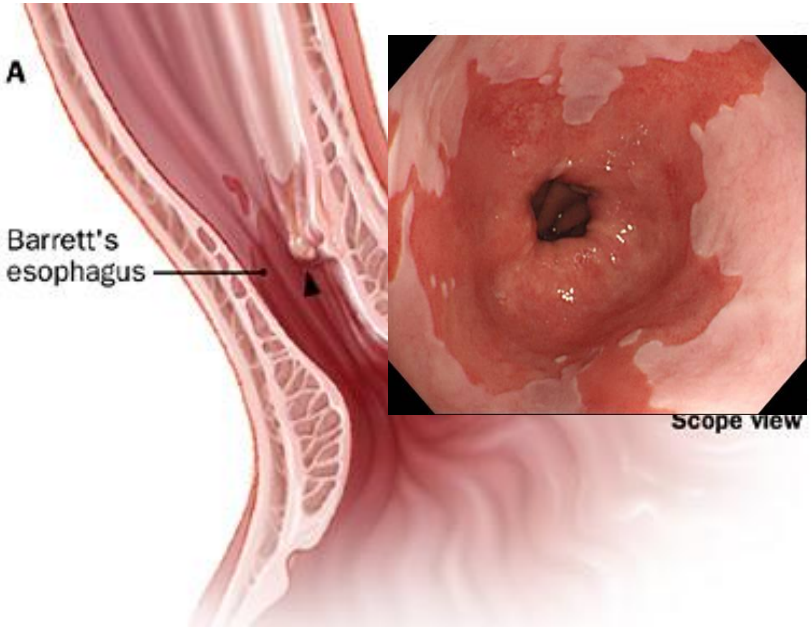
03. XAI

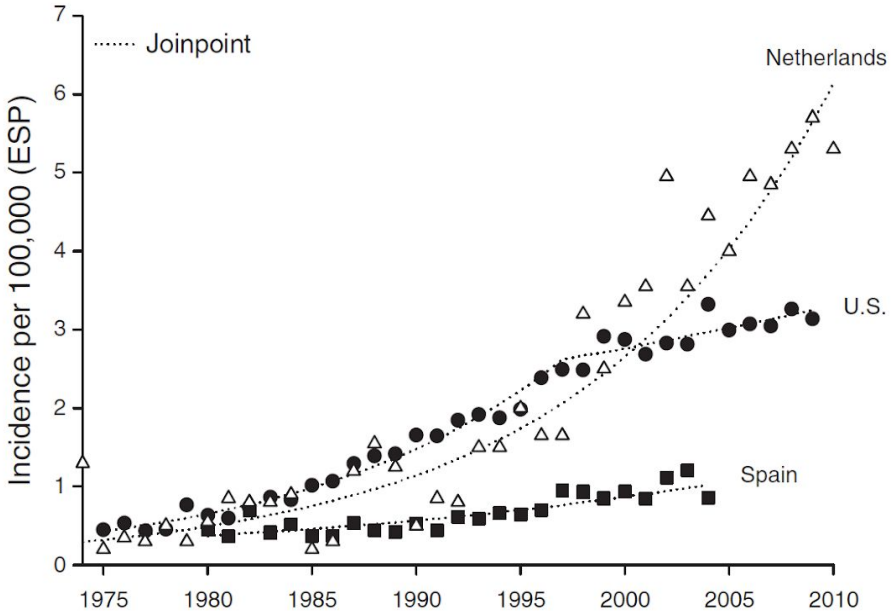
04. Discussion

Case: Barrett's Oesophagus

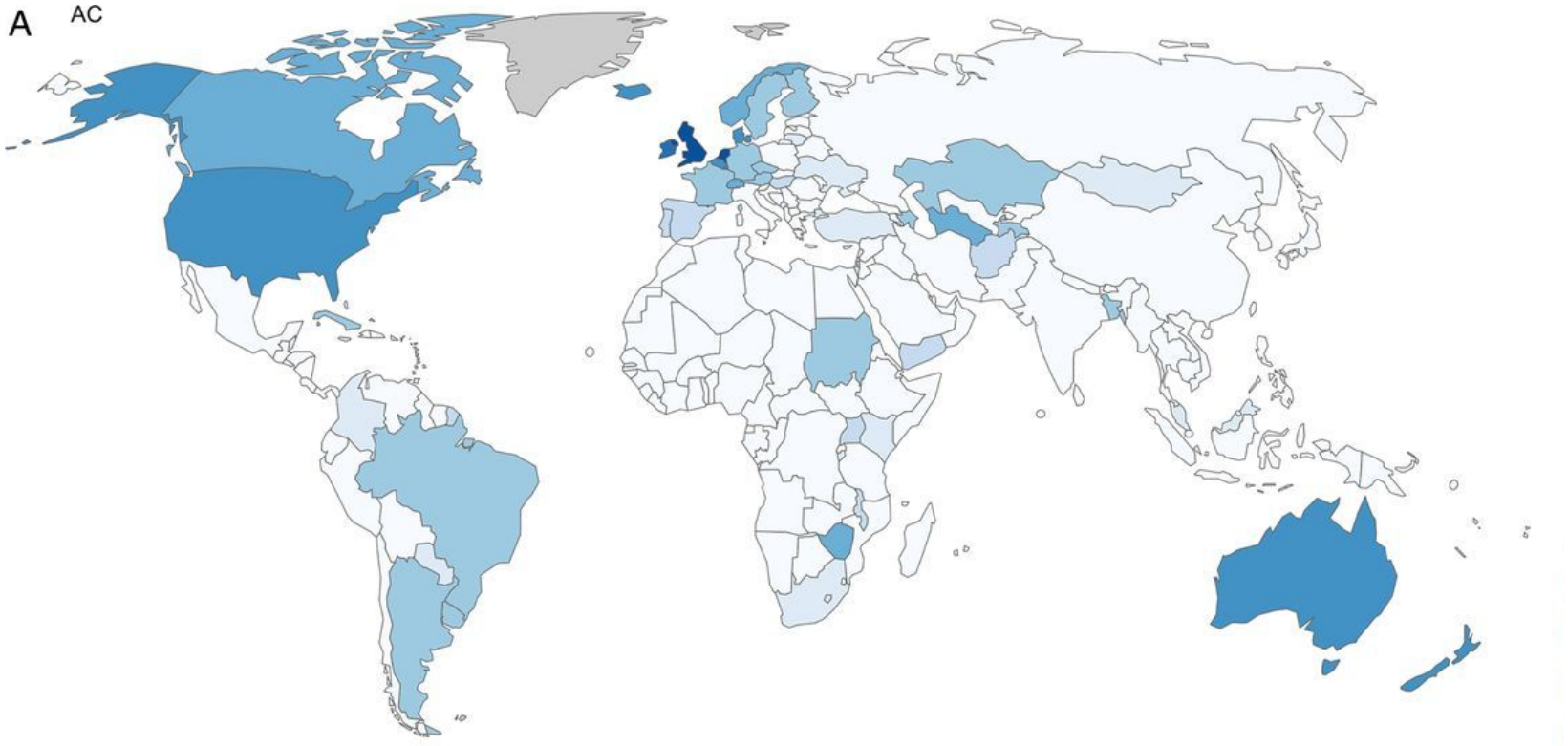


Case: Barrett's Oesophagus

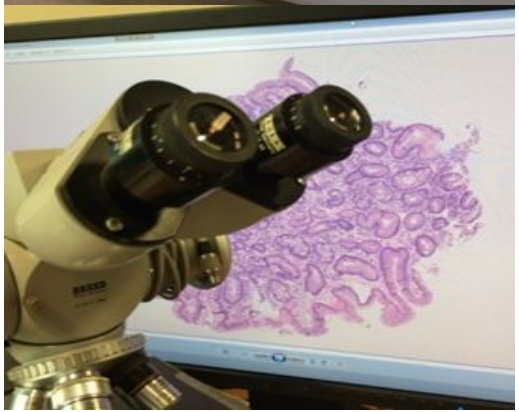


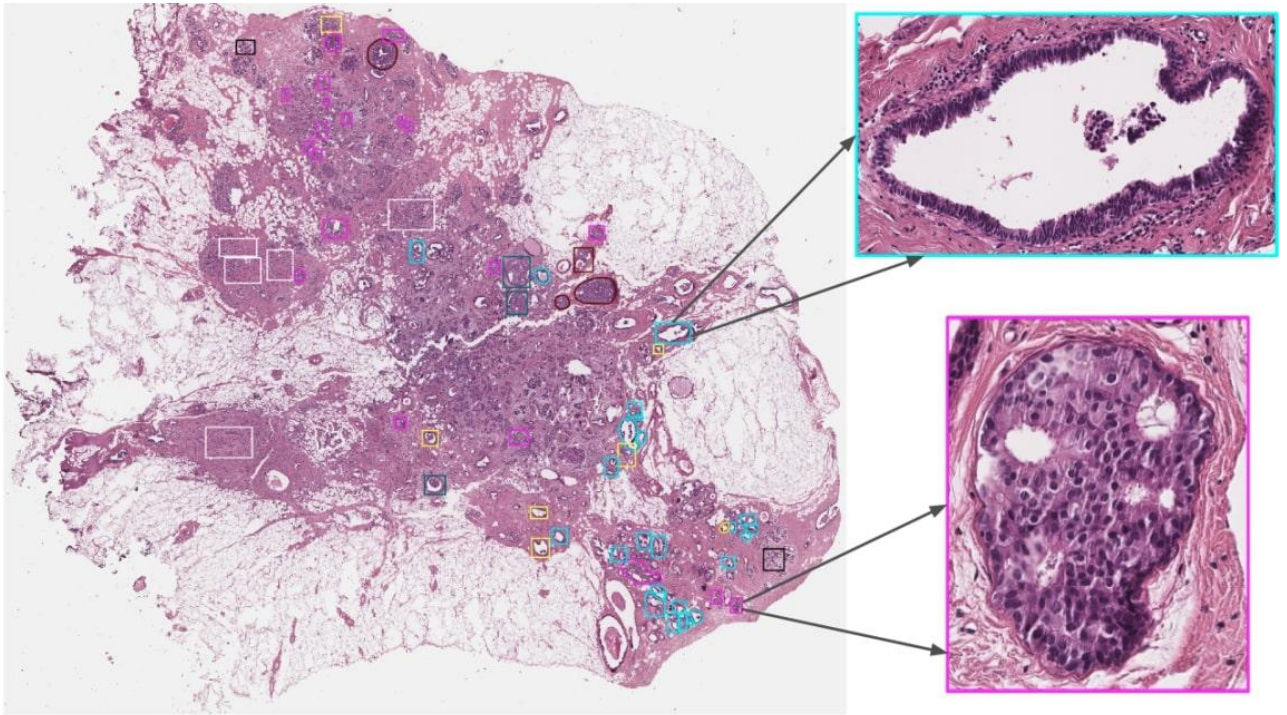


Case: Barrett's Oesophagus

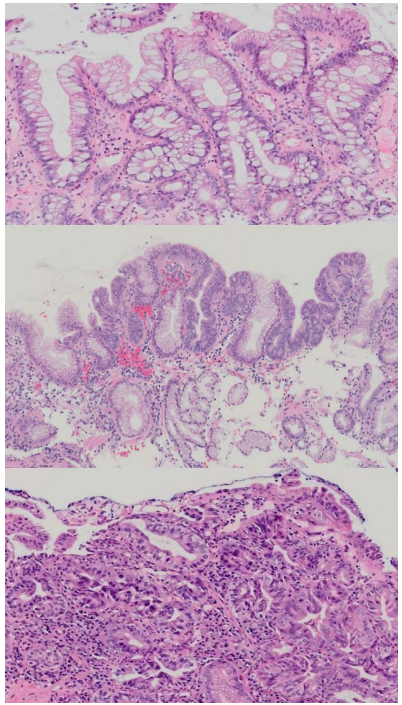


Case: Barrett's Oesophagus





Barrett's Oesophagus Spectrum



01. **Non Dysplastic Barrett's**
Low Encoscopic Surveillance

02. **Low Grade Dysplasia**
Encoscopic Ablation Therapy

03. **High Grade Dysplasia**
Surgical Therapy

A proportion of patients develop Adenocarcinoma

Risk of progression of Non-Dysplastic Barrett's Epitheleum to cancer is low:

~0.3 - 0.6 % per year

CARCINOMA

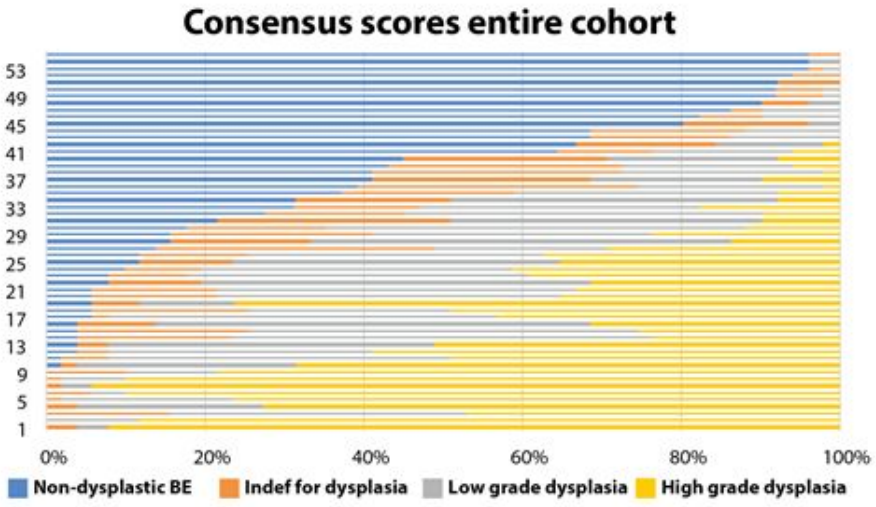
SURF

Interobserver Variability in Histopathology

Gastroenterology 2017;152:564–570

Discordance Among Pathologists in the United States and Europe in Diagnosis of Low-Grade Dysplasia for Patients With Barrett's Esophagus

Prashanth Vennalaganti,[1,2] Vijay Kanakadandi,[1,2] John R. Goldblum,[3] Sharad C. Mathur,[4] Deepa T. Patil,[3] G. Johan Offerhaus,[5] Sybren L. Meijer,[6] Michael Vieth,[7] Robert D. Odze,[8] Saligram Shreyas,[1,2] Sravanthi Parasa,[1,2] Neil Gupta,[9] Alessandro Repici,[10] Ajay Bansal,[1,2] Titi Mohammad,[1,2] and Prateek Sharma[1,2]



American Journal of Gastroenterology
© 2008 by Am. Coll. of Gastroenterology
Published by Blackwell Publishing
ISSN 0002-9270
doi: 10.1111/j.1572-0241.2008.02020.x

ORIGINAL CONTRIBUTIONS

CME

Pathology

Poor Interobserver Agreement in the Distinction of High-Grade Dysplasia and Adenocarcinoma in Pretreatment Barrett's Esophagus Biopsies

Erinn Downs-Kelly, D.O.,[1] Joel E. Mendelin, M.D.,[1] Ana E. Bennett, M.D.,[1] Elias Castilla, M.D.,[1] Walter H. Henricks,[1] Lynn Schoenfield, M.D.,[1] Marek Skacel, M.D.,[1] Lisa Yerian, M.D.,[1] Thomas W. Rice, M.D.,[2] Lisa A. Rybicki, M.S.,[3] Mary P. Bronner, M.D.,[1] and John R. Goldblum, M.D.[1]
[1]*Cleveland Clinic Departments of Anatomic Pathology,* [2]*Thoracic Surgery, and* [3]*Quantitative Health Sciences, Cleveland, Ohio*

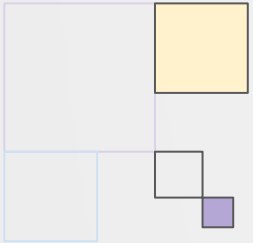
Interobserver Study of 55 pathologists

Variation in diagnostic accuracy of 46%-97% (median 66%)

Desire for **Expert Review**

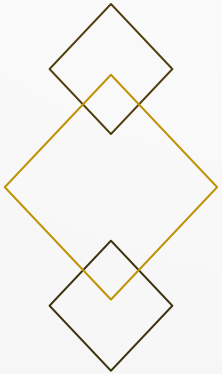
SURF

09/11/2022



Data & Deep Learning

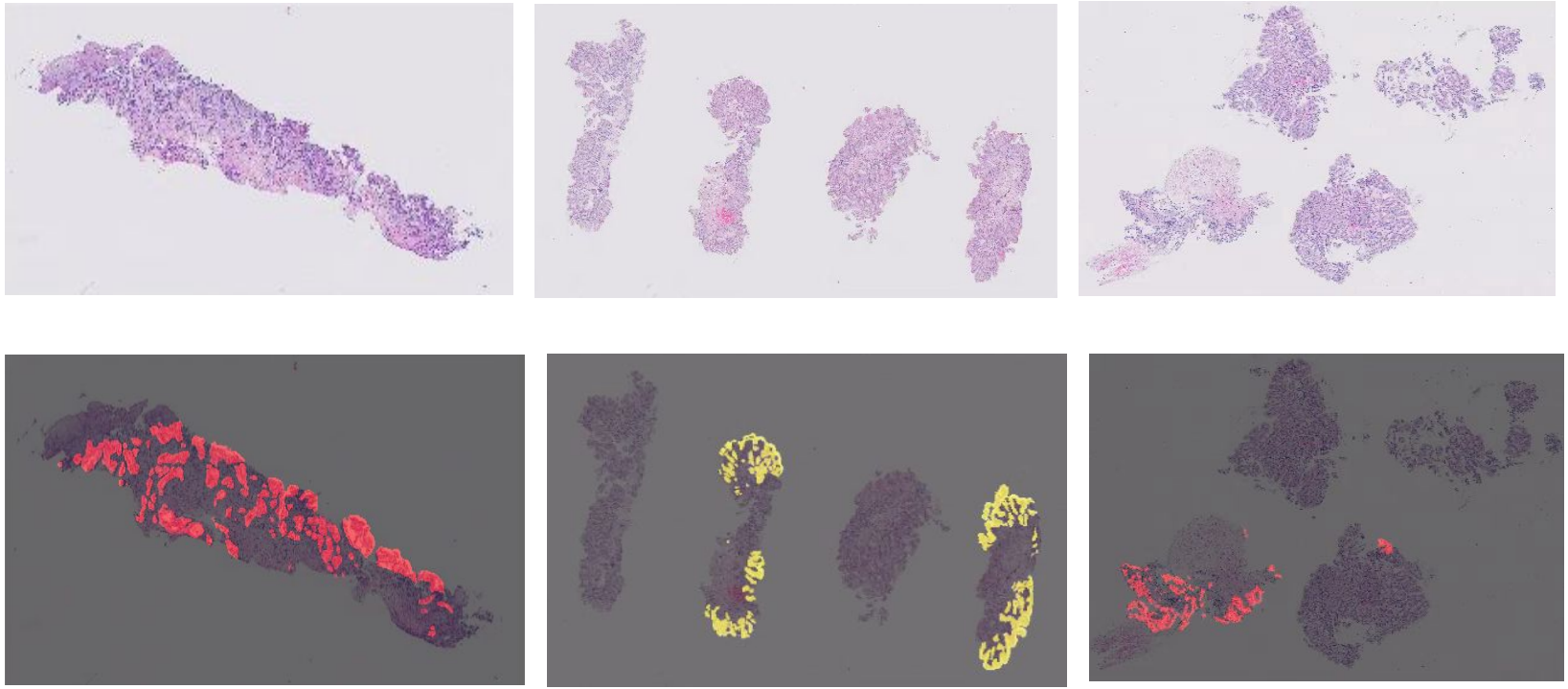
A segmentation case



High Performance Machine Learning Group

SURF

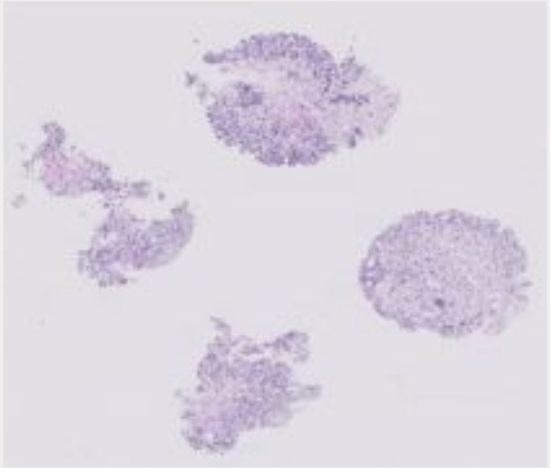
Histopathology Segmentation

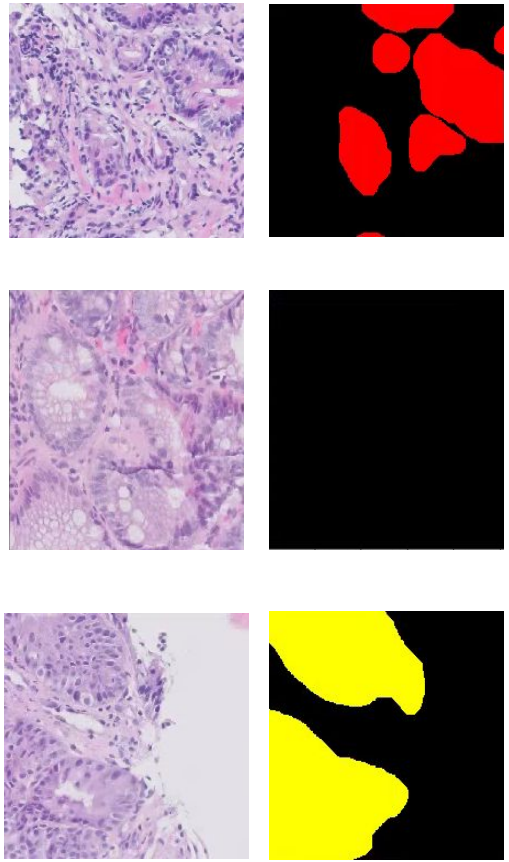


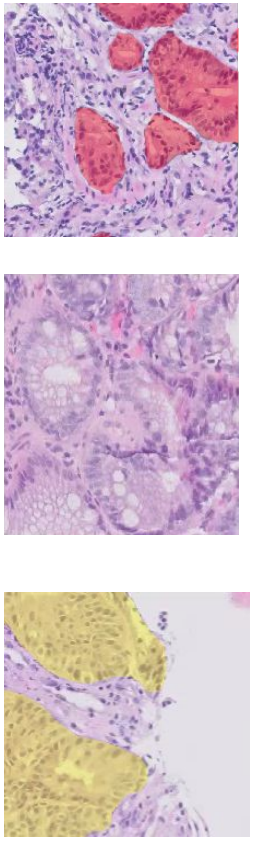
Slice up the whole slide images (WSIs)

Slicing usually results in **thousands** of smaller images per WSI

Slices of 256x256 or 512x512







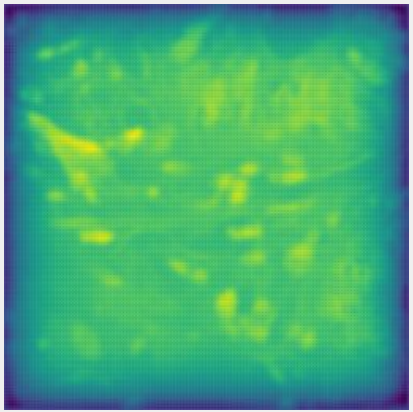
SURF

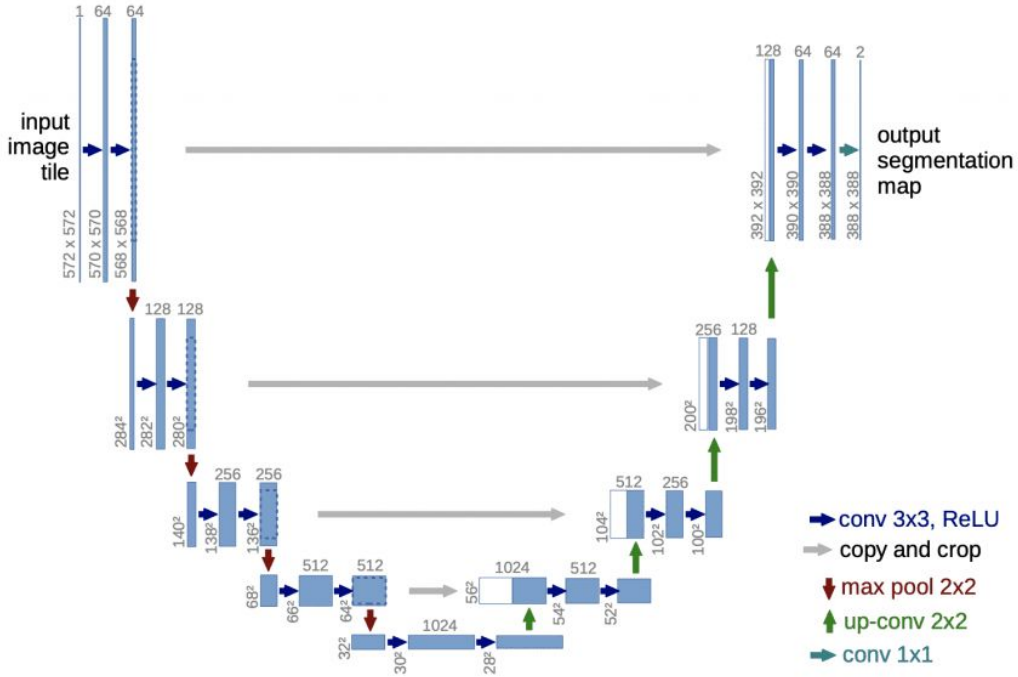
U-Net: Master of Segmentation

U-Net is a fundamental architecture in segmentation

Encoder and decoder structure

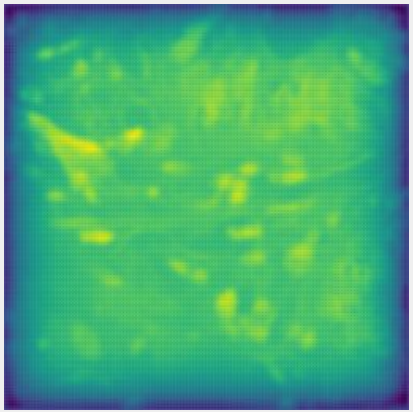
Skip-Connections (from ResNets)

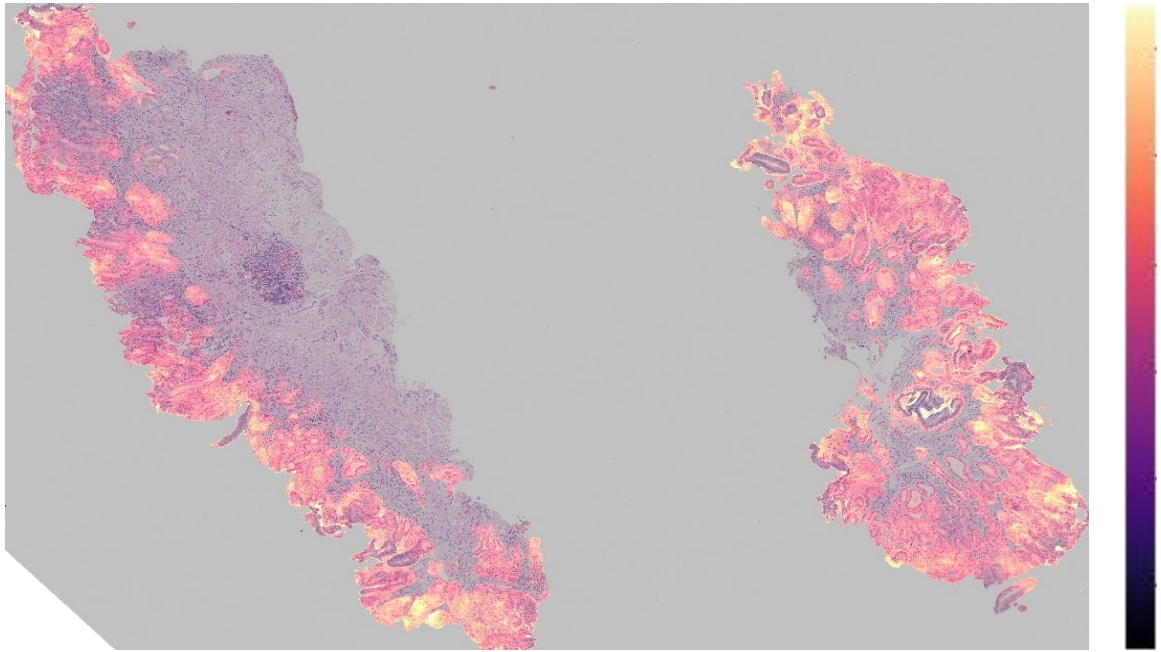




U-Net: Master of Segmentation

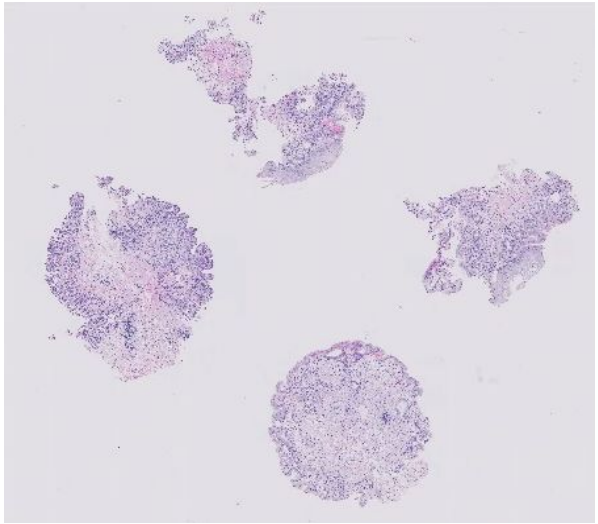
U-Net outputs softmax distribution



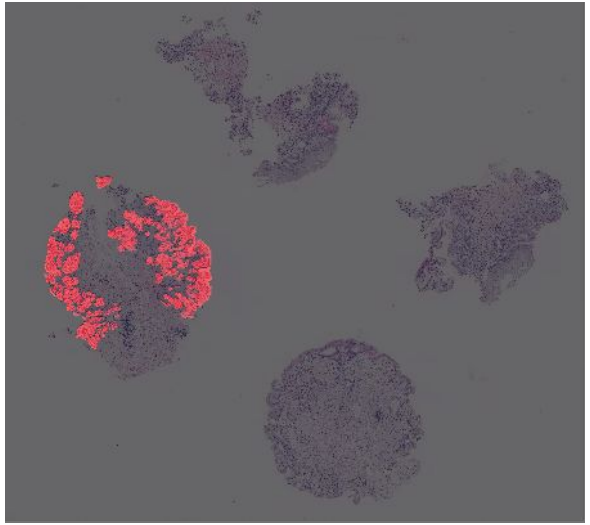


Example Multi Class Predictions

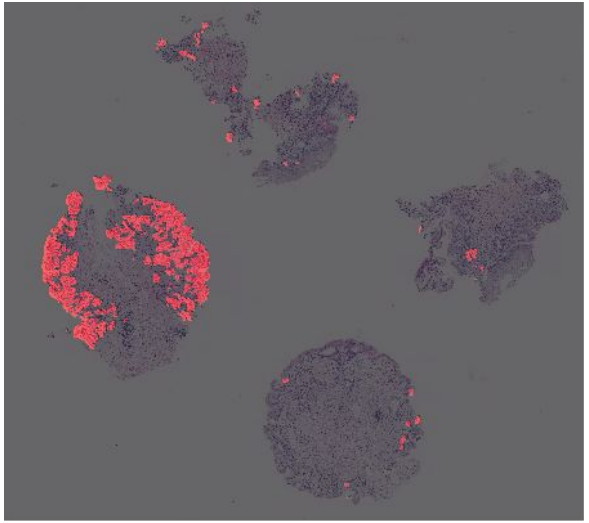
Tissue WSI



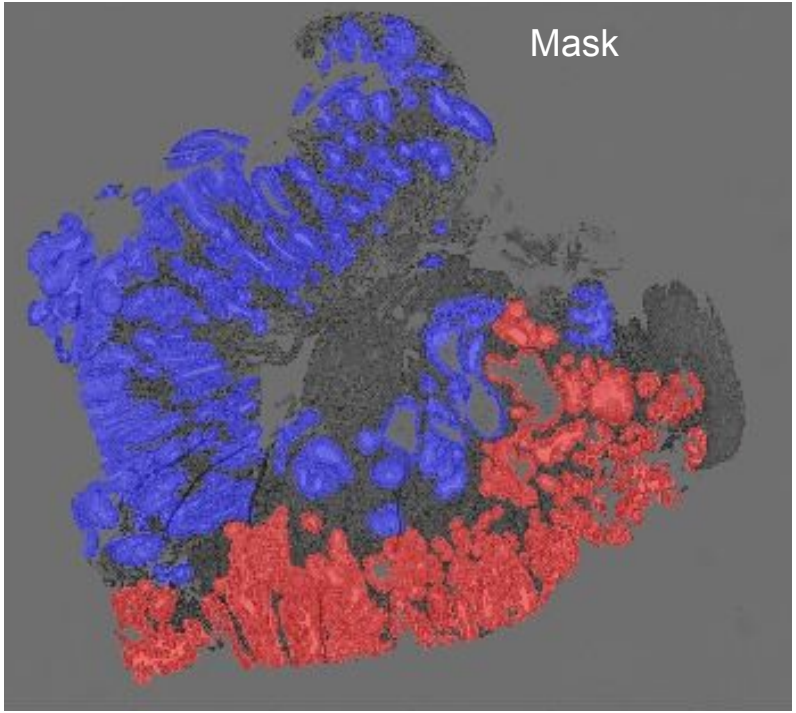
Annotated Mask

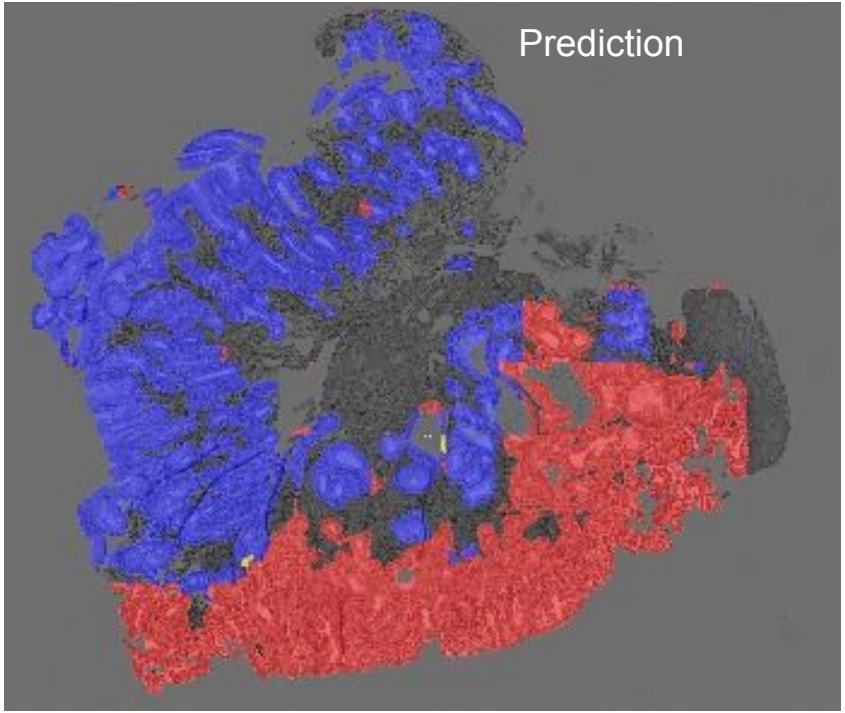


Predicted Mask

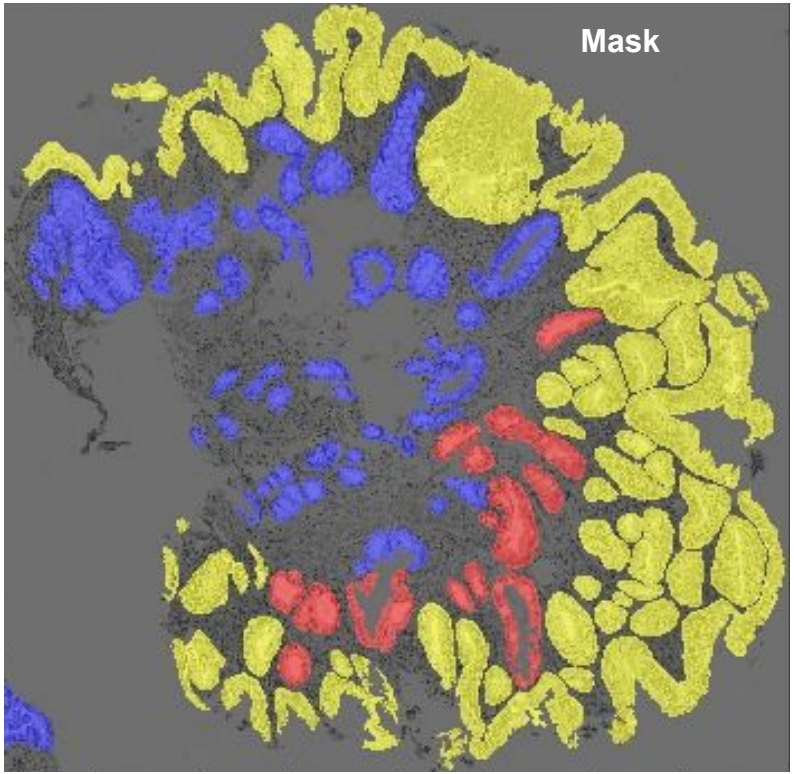


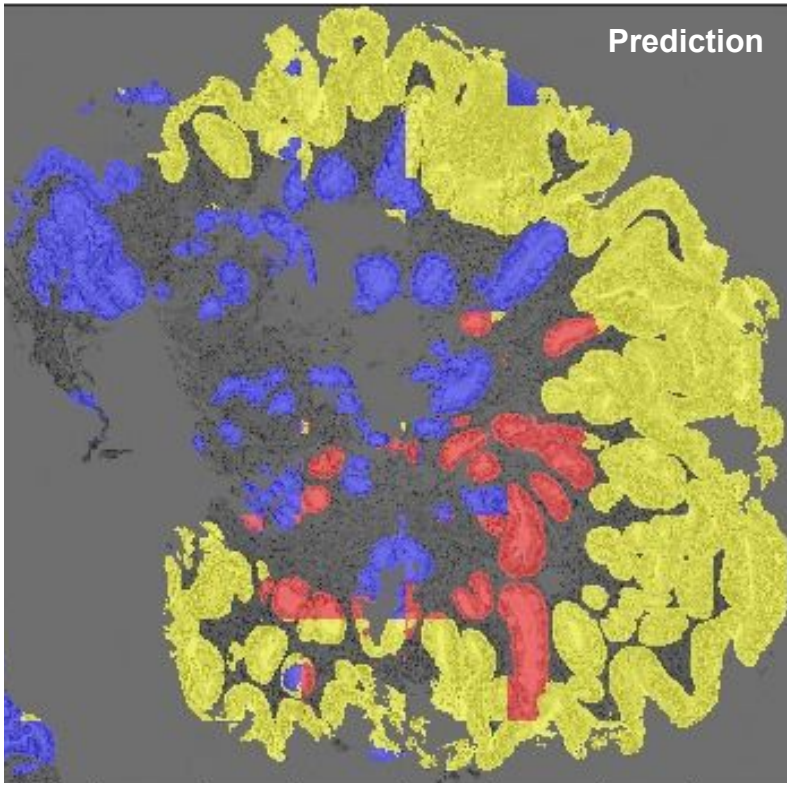
Example Multi Class Predictions



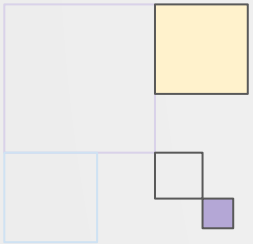


Example Multi Class Predictions



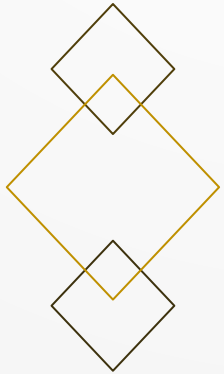


09/11/2022



CLIP: Representations

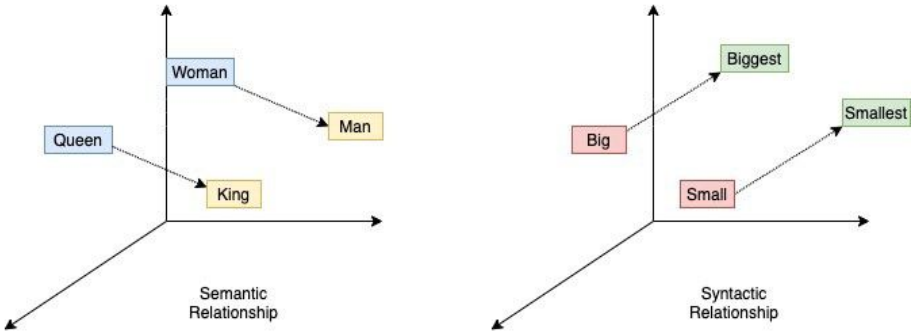
Connecting images with texts

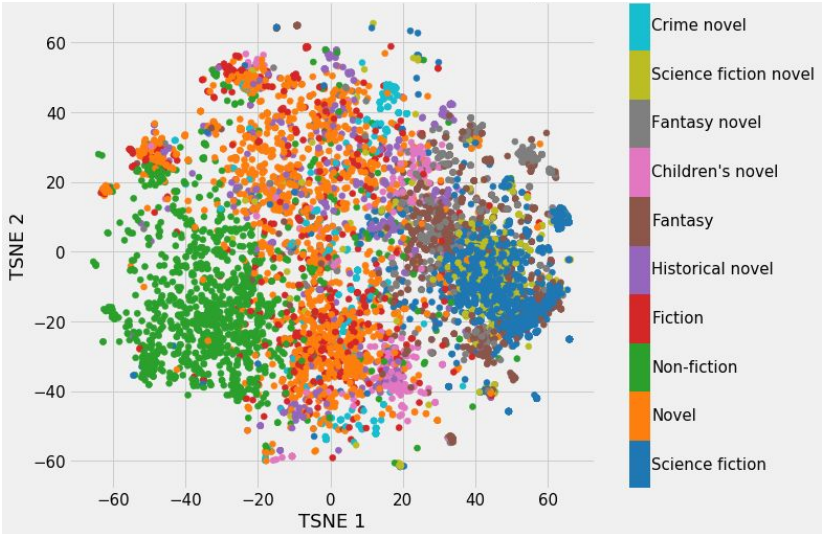


High Performance Machine Learning Group

SURF

Neural Representation Learning

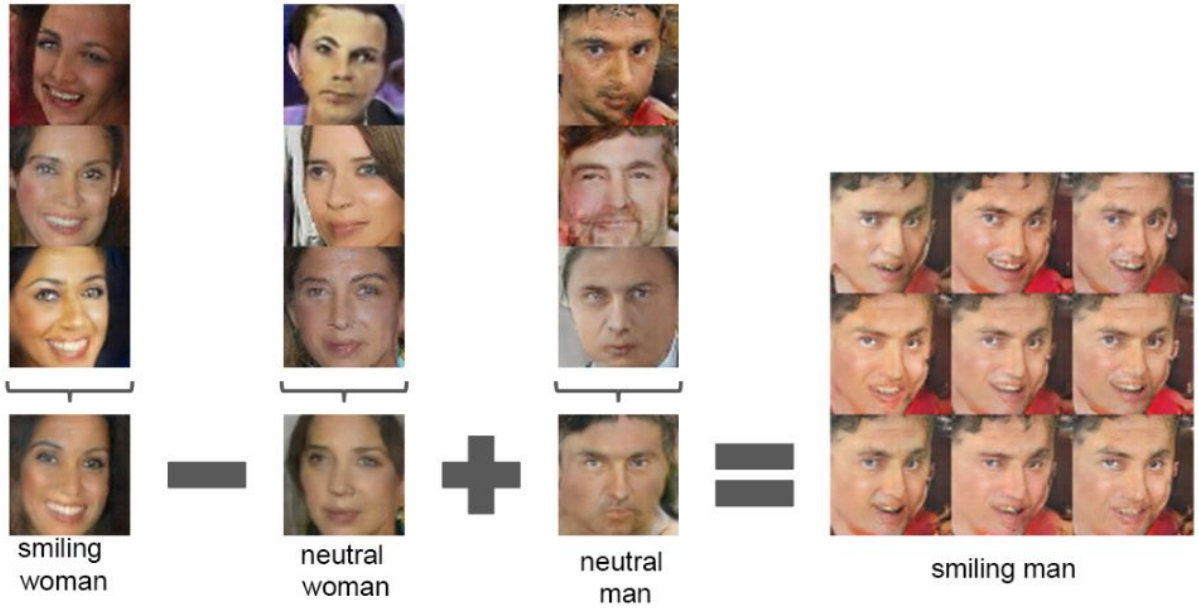




King - Man + Woman = Queen

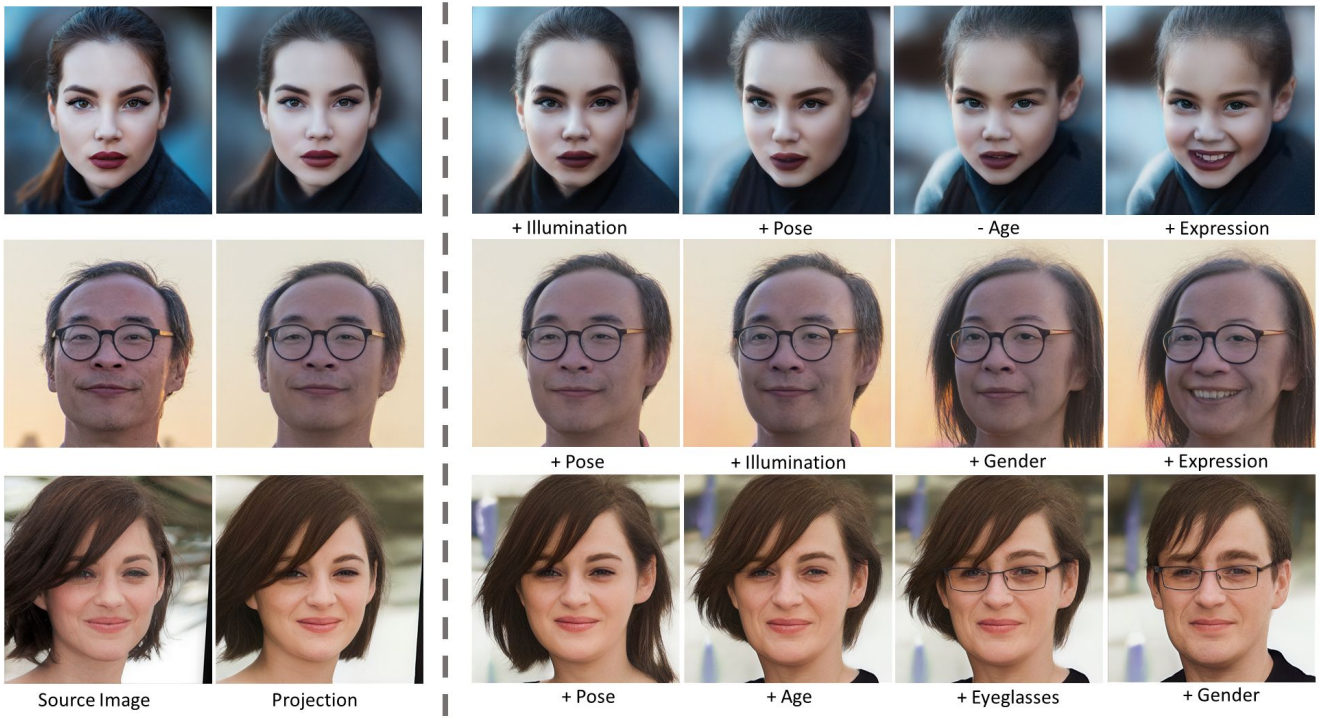
Each word has a numerical vector representation learnt by a language model

Neural Representation Learning



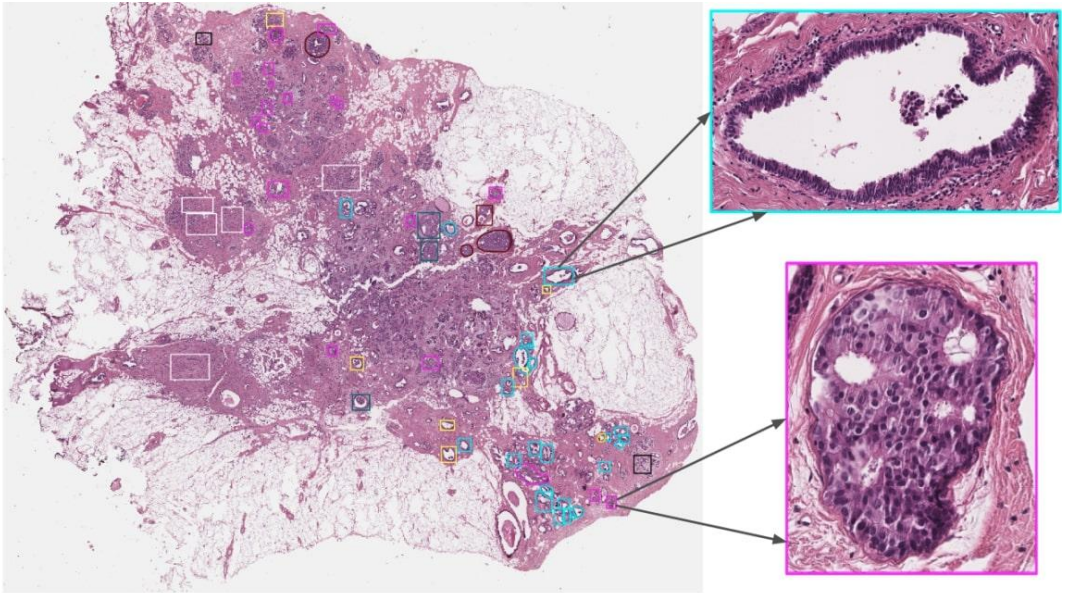
Mirza & Osindero (2014, arXiv:1411.1784); Isola et al. (2016, arXiv:1611.07004)

Neural Representation Learning



"StyleFlow: Attribute-conditioned Exploration of StyleGAN-Generated Images using Conditional Continuous Normaliz

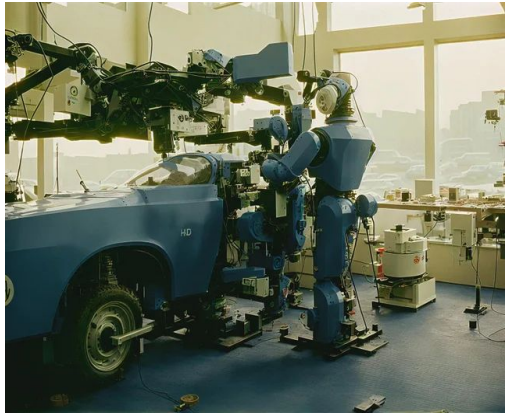
Examode Data

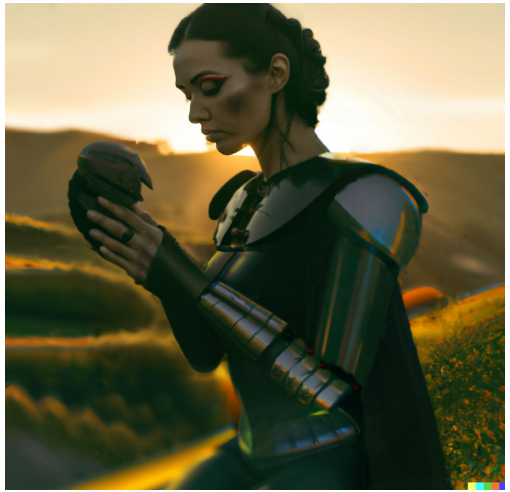


- “colon biopsy at 21-25 cm and at 15 cm: in both subheadings metastasis of adenocarcinoma, morphological and immune histochemicalally best suited to metastasis of the patient's known primary ovarian carcinoma.”
- 2 - tubular adenoma with low dysplasia and, focally, high degree (moderate, focally severe).

The Engine of DALL-E

- Trained on 500M+ image and text pairs
- Stability AI used 4000 A100 GPUs
- Using CLIP to obtain good text <-> image representations





The Engine of DALL-E

- Trained on 500M+ image and text pairs
- Stability AI used 4000 A100 GPUs
- Using CLIP to obtain good text <-> image representations

A photo of a dog

Raccoon climbing a tree

The Engine of DALL-E

- Trained on 500M+ image and text pairs
- Stability AI used 4000 A100 GPUs
- Using CLIP to obtain good text <-> image representations

✓ a photo of **guacamole**, a type of food.

× a photo of **ceviche**, a type of food.

× a photo of **edamame**, a type of food.

× a photo of **tuna tartare**, a type of food.

× a photo of **hummus**, a type of food.

✓ a photo of a **television studio**.

× a photo of a **podium indoor**.

× a photo of a **conference room**.

× a photo of a **lecture room**.

× a photo of a **control room**.

SURF

The Engine of DALL-E

- Trained on 500M+ image and text pairs
- Stability AI used 4000 A100 GPUs
- Using CLIP to obtain good text <-> image representations

Use **CLIP** to find specific visuals

The Engine of DALL-E

- Trained on 500M+ image and text pairs
- Stability AI used 4000 A100 GPUs
- Using CLIP to obtain good text <-> image representations

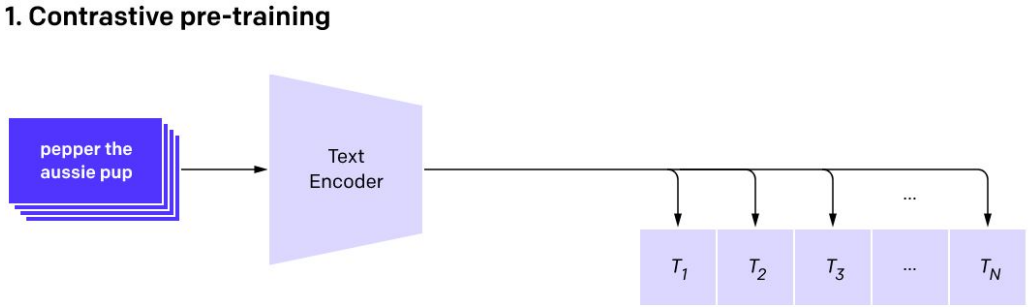
A bicyclist with a blue shirt

The Engine of DALL-E

- Trained on 500M+ image and text pairs
- Stability AI used 4000 A100 GPUs
- Using CLIP to obtain good text <-> image representations

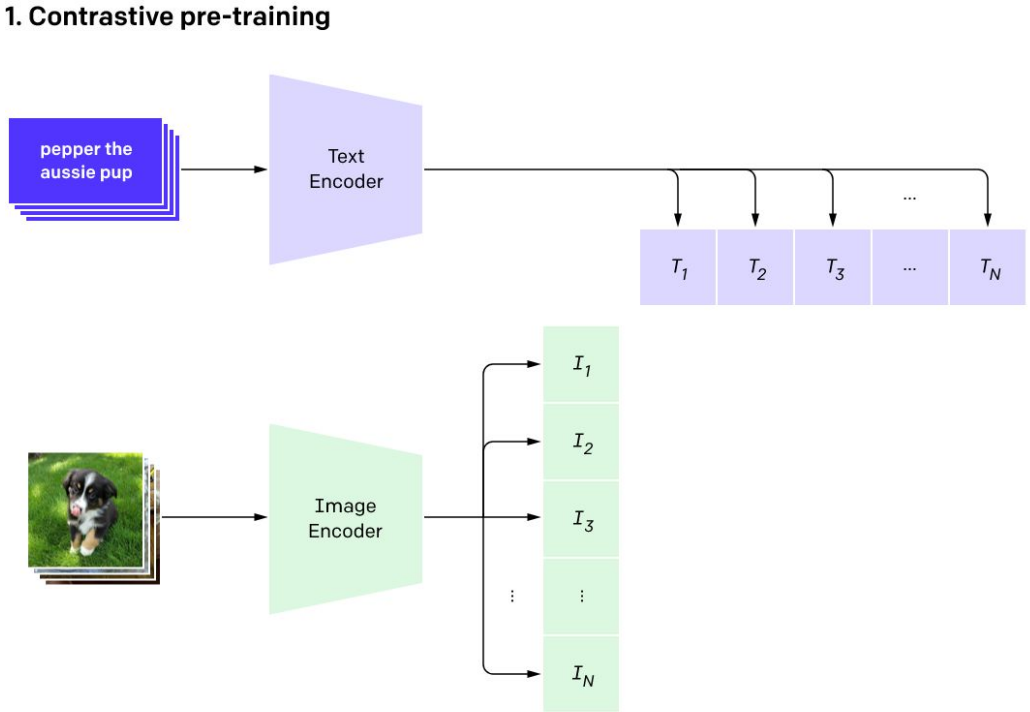
A truck with the text “JCN”

Contrastive Language-Image Pre-training



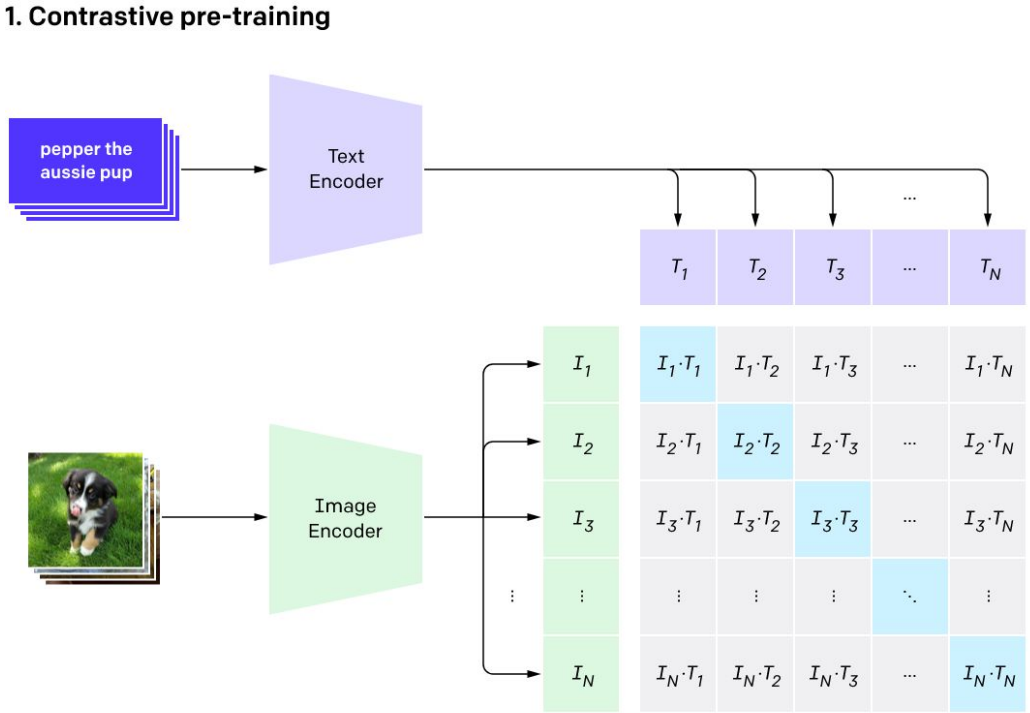
https://openai.com/blog/clip/

Contrastive Language-Image Pre-training



https://openai.com/blog/clip/

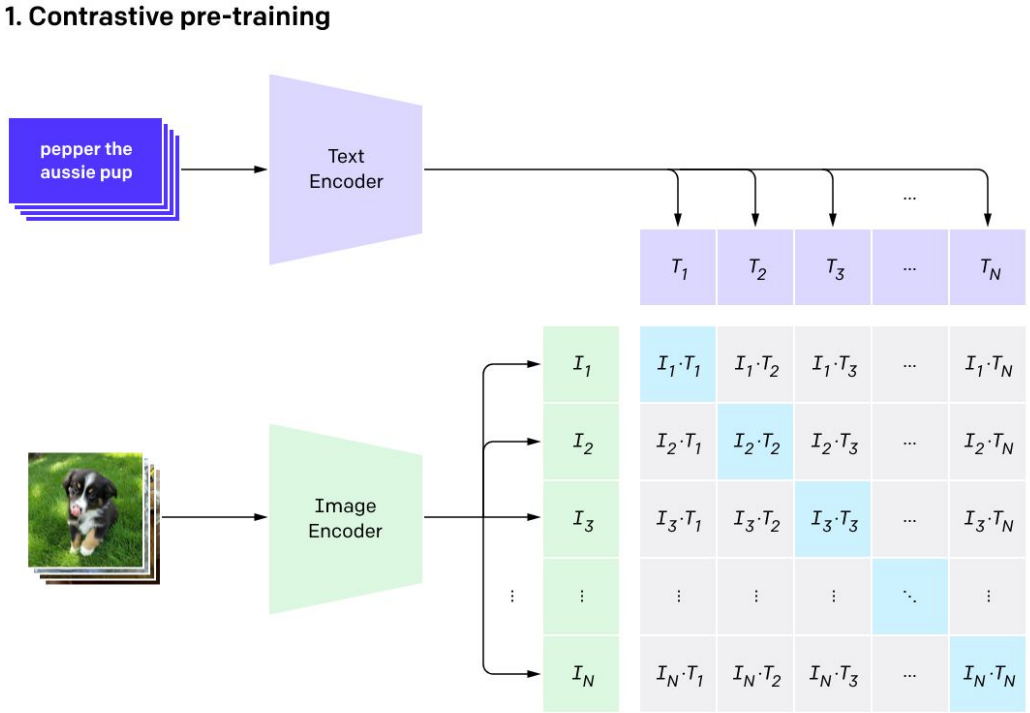
Contrastive Language-Image Pre-training



Dot product between text encoder vector and image encoder vector

https://openai.com/blog/clip/

Contrastive Language-Image Pre-training

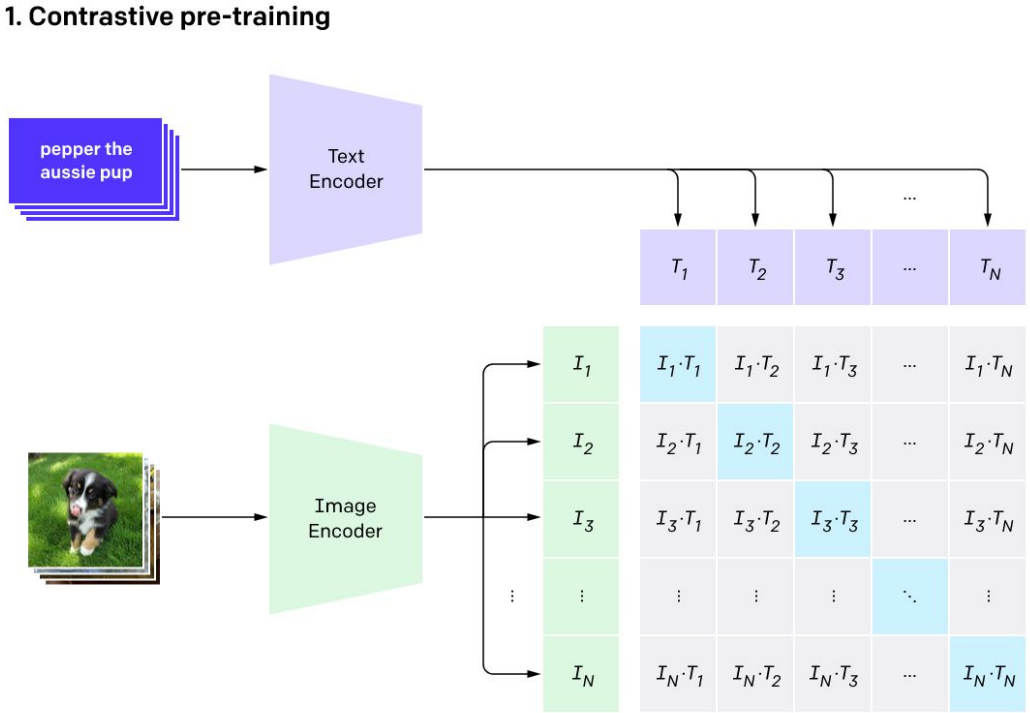


Dot product between text encoder vector and image encoder vector

Batches: N (image, text) pairs

https://openai.com/blog/clip/

Contrastive Language-Image Pre-training



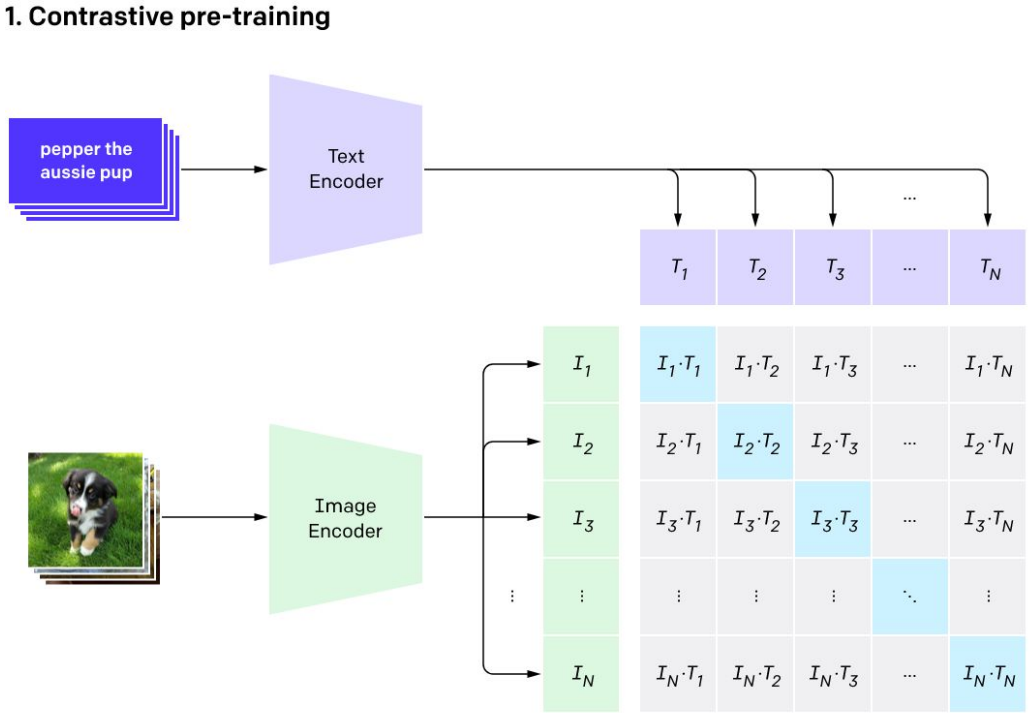
Dot product between text encoder vector and image encoder vector

Batches: N (image, text) pairs

Trained to predict which of the N × N possible pairings occur **across a batch**

https://openai.com/blog/clip/

Contrastive Language-Image Pre-training



Dot product between text encoder vector and image encoder vector

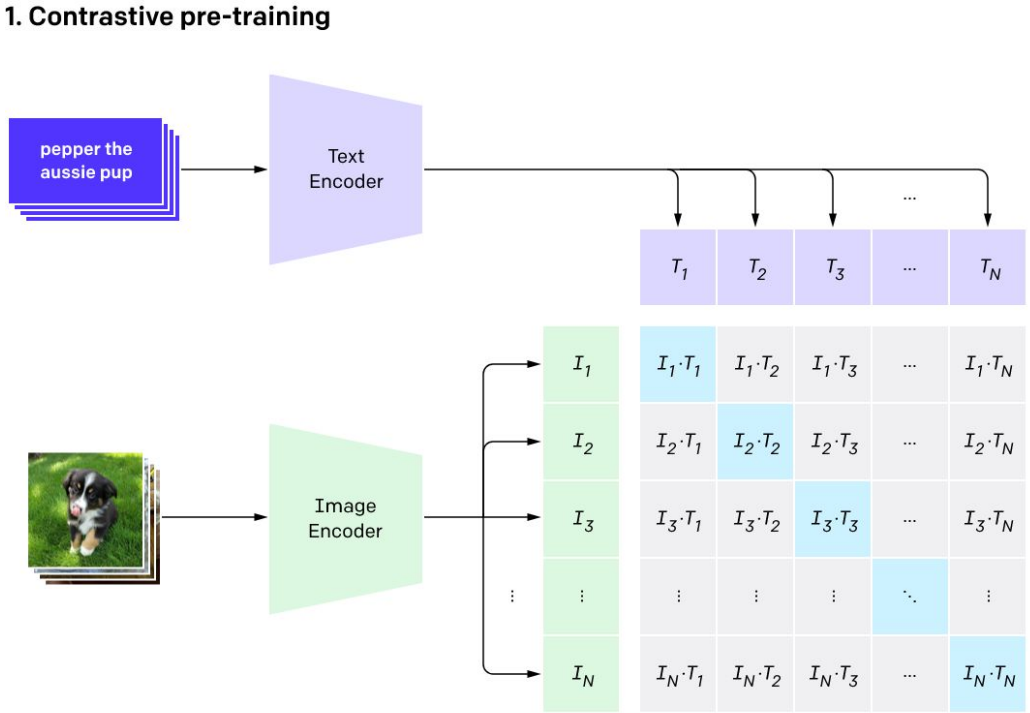
Batches: N (image, text) pairs

Trained to predict which of the N × N possible pairings occur **across a batch**

Bigger batch size → Better representations

https://openai.com/blog/clip/

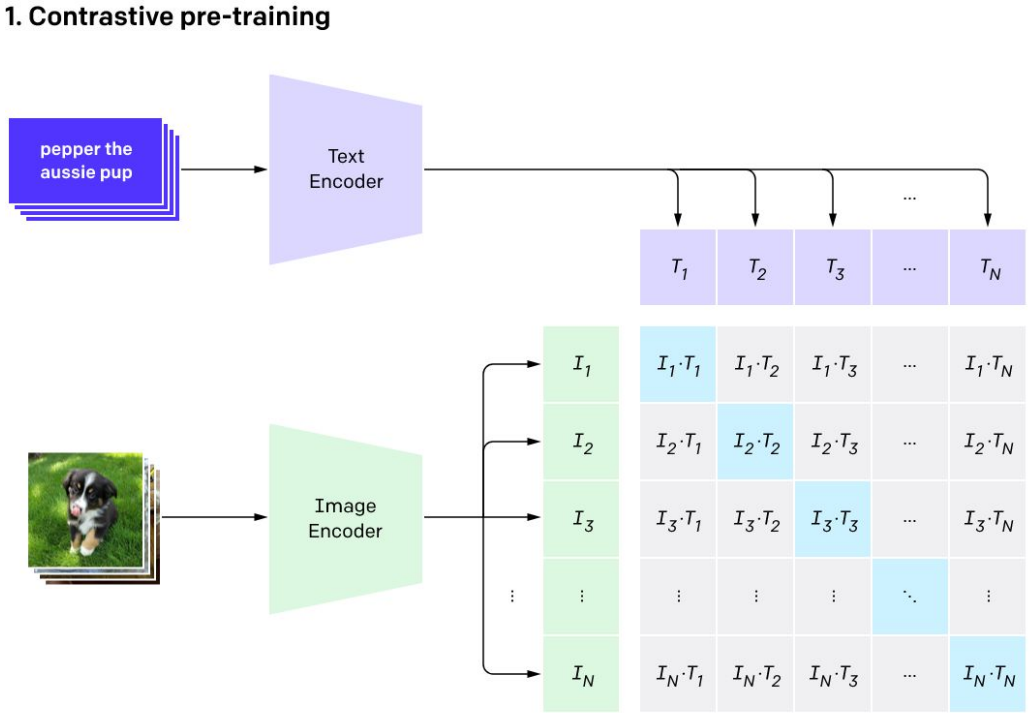
Contrastive Language-Image Pre-training




```
I_f = image_encoder(I)
T_f = image_encoder(T)

I_e = l2_normalize(dot(I_f, W_i))
T_e = l2_normalize(dot(T_f, W_t))

output = dot(I_e, T_e.T) * exp(t)

labels = arange(n)
loss_i = CE(output, labels, axis=0)
loss_t = CE(output, labels, axis=1)

loss = (loss_i + loss_t)/2
```

https://openai.com/blog/clip/

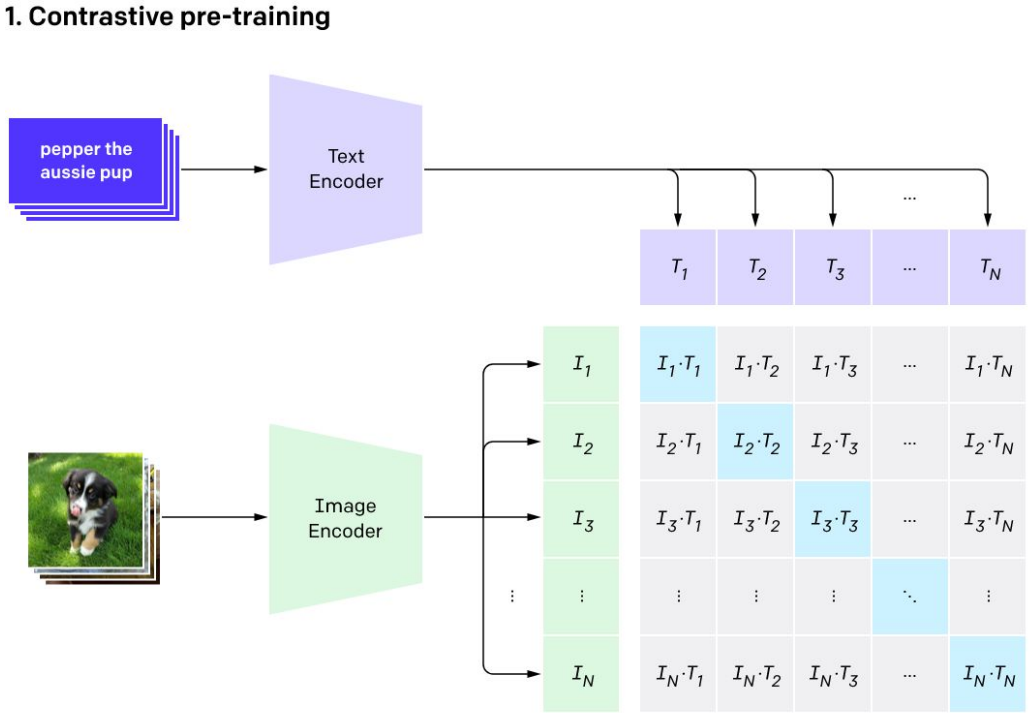
Contrastive Language-Image Pre-training




```
I_f = image_encoder(I)
T_f = image_encoder(T)

I_e = l2_normalize(dot(I_f, W_i))
T_e = l2_normalize(dot(T_f, W_t))

output = dot(I_e, T_e.T) * exp(t)

labels = arange(n)
loss_i = CE(output, labels, axis=0)
loss_t = CE(output, labels, axis=1)

loss = (loss_i + loss_t)/2
```

https://openai.com/blog/clip/

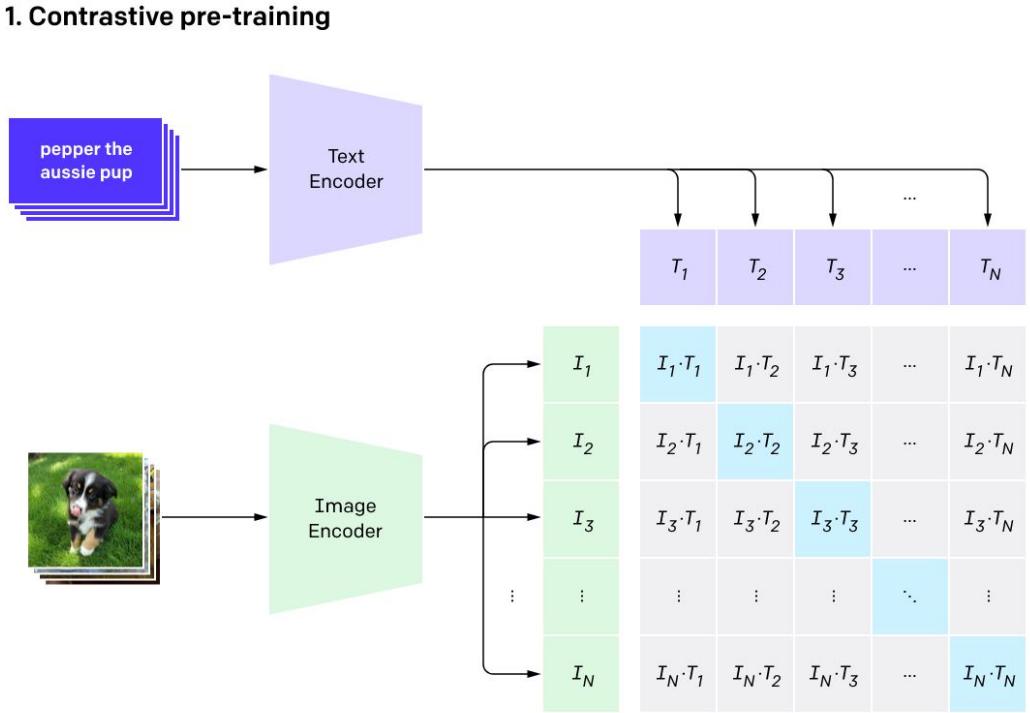
Contrastive Language-Image Pre-training




```
I_f = image_encoder(I)
T_f = image_encoder(T)

I_e = l2_normalize(dot(I_f, W_i))
T_e = l2_normalize(dot(T_f, W_t))

output = dot(I_e, T_e.T) * exp(t)

labels = arange(n)
loss_i = CE(output, labels, axis=0)
loss_t = CE(output, labels, axis=1)

loss = (loss_i + loss_t)/2
```

https://openai.com/blog/clip/

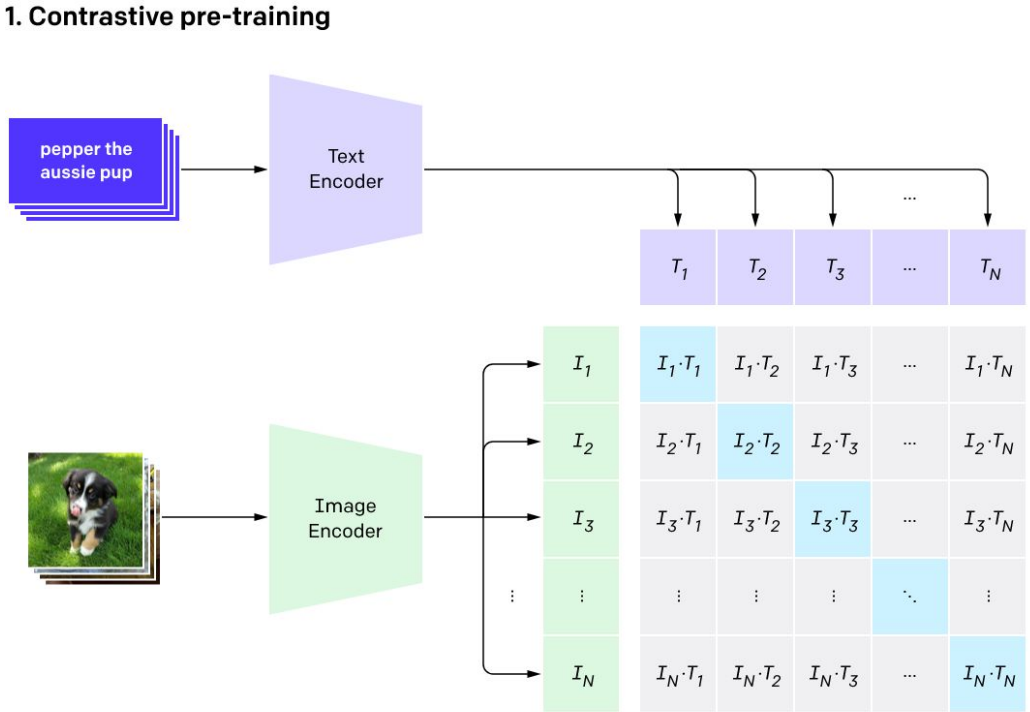
Contrastive Language-Image Pre-training




```
I_f = image_encoder(I)
T_f = image_encoder(T)

I_e = l2_normalize(dot(I_f, W_i))
T_e = l2_normalize(dot(T_f, W_t))

output = dot(I_e, T_e.T) * exp(t)

labels = arange(n)
loss_i = CE(output, labels, axis=0)
loss_t = CE(output, labels, axis=1)

loss = (loss_i + loss_t)/2
```

https://openai.com/blog/clip/

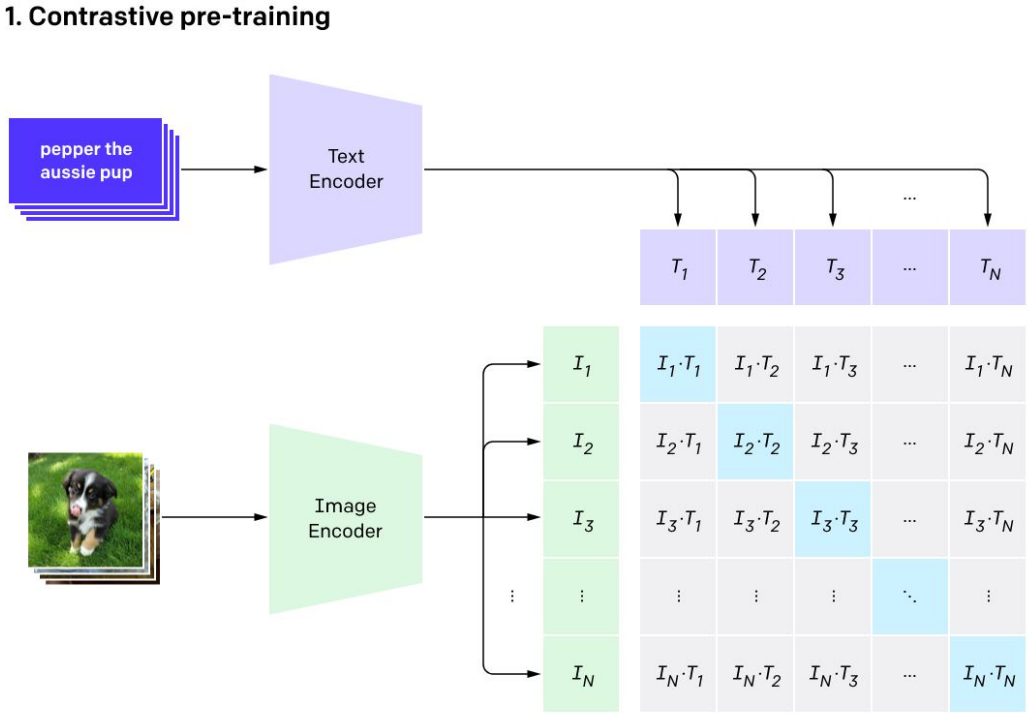
Contrastive Language-Image Pre-training




```
I_f = image_encoder(I)
T_f = image_encoder(T)

I_e = l2_normalize(dot(I_f, W_i))
T_e = l2_normalize(dot(T_f, W_t))

output = dot(I_e, T_e.T) * exp(t)

labels = arange(n)
loss_i = CE(output, labels, axis=0)
loss_t = CE(output, labels, axis=1)

loss = (loss_i + loss_t)/2
```

https://openai.com/blog/clip/

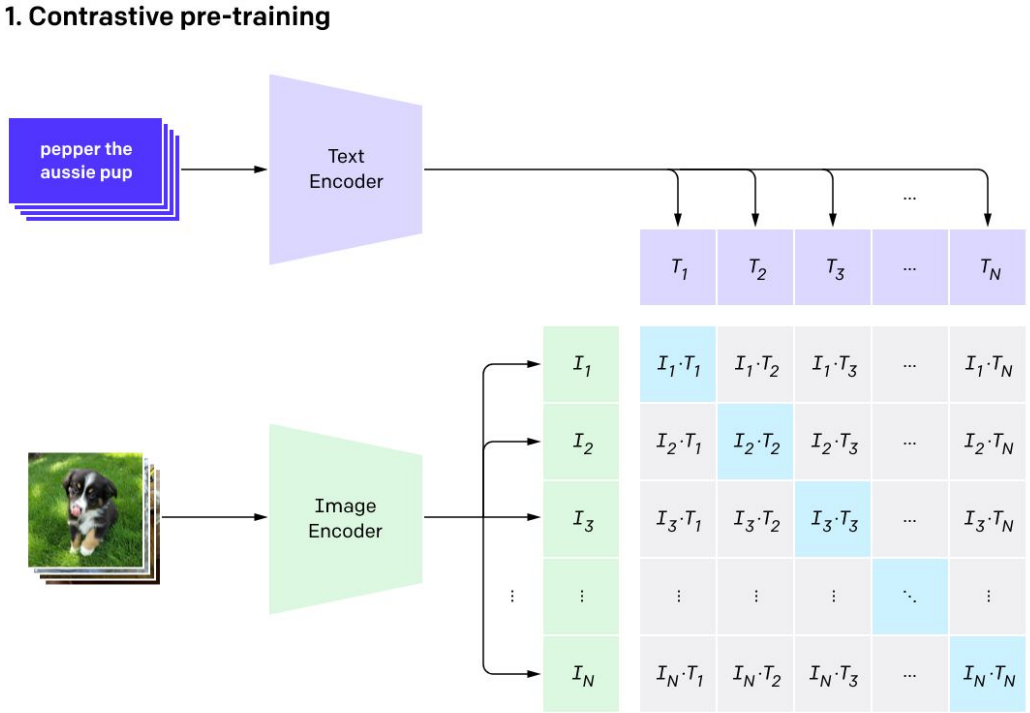
Contrastive Language-Image Pre-training




```
I_f = image_encoder(I)
T_f = image_encoder(T)

I_e = l2_normalize(dot(I_f, W_i))
T_e = l2_normalize(dot(T_f, W_t))

output = dot(I_e, T_e.T) * exp(t)

labels = arange(n)
loss_i = CE(output, labels, axis=0)
loss_t = CE(output, labels, axis=1)

loss = (loss_i + loss_t)/2
```

https://openai.com/blog/clip/

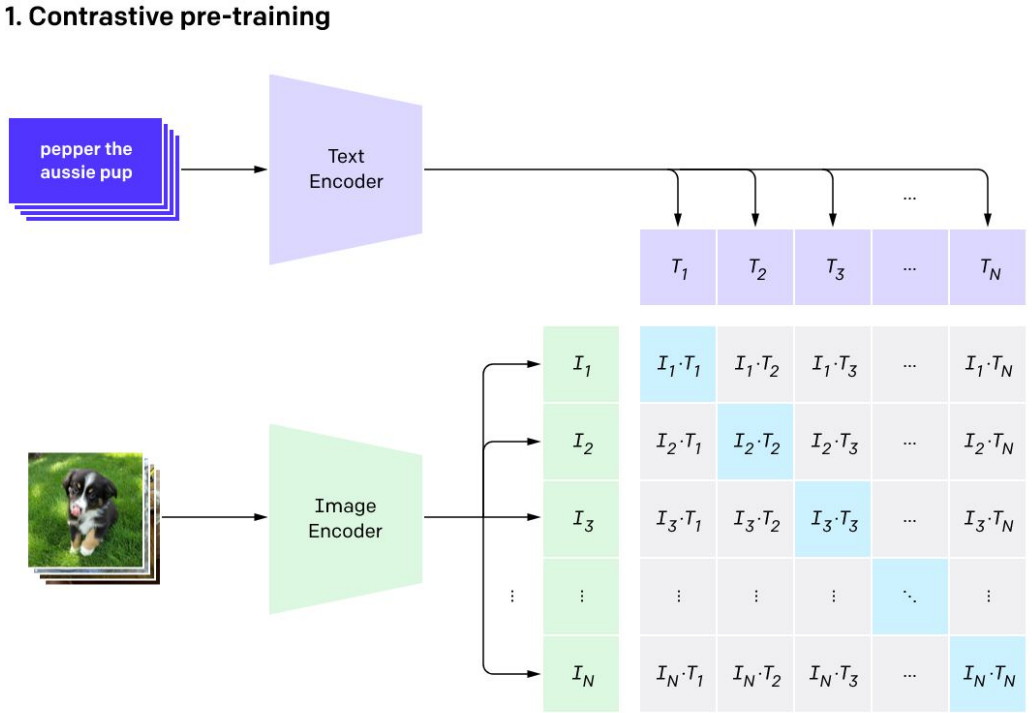
Contrastive Language-Image Pre-training




```
I_f = image_encoder(I)
T_f = image_encoder(T)

I_e = l2_normalize(dot(I_f, W_i))
T_e = l2_normalize(dot(T_f, W_t))

output = dot(I_e, T_e.T) * exp(t)

labels = arange(n)
loss_i = CE(output, labels, axis=0)
loss_t = CE(output, labels, axis=1)

loss = (loss_i + loss_t)/2
```

https://openai.com/blog/clip/

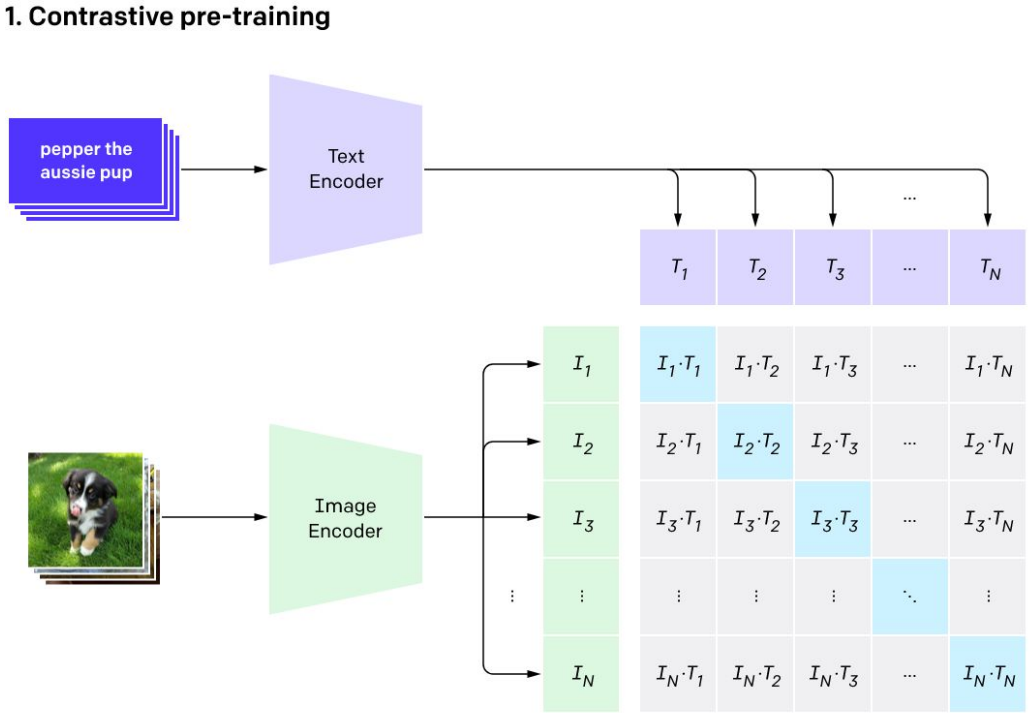
Contrastive Language-Image Pre-training




```
I_f = image_encoder(I)
T_f = image_encoder(T)

I_e = l2_normalize(dot(I_f, W_i))
T_e = l2_normalize(dot(T_f, W_t))

output = dot(I_e, T_e.T) * exp(t)

labels = arange(n)
loss_i = CE(output, labels, axis=0)
loss_t = CE(output, labels, axis=1)

loss = (loss_i + loss_t)/2
```

https://openai.com/blog/clip/

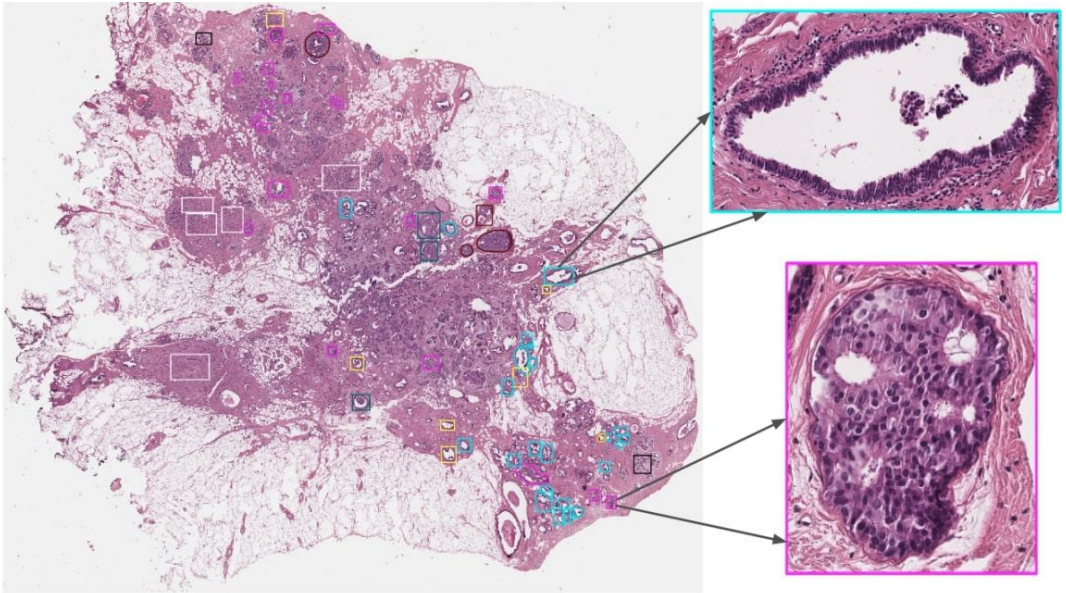
Contrastive Language-Image Pre-training



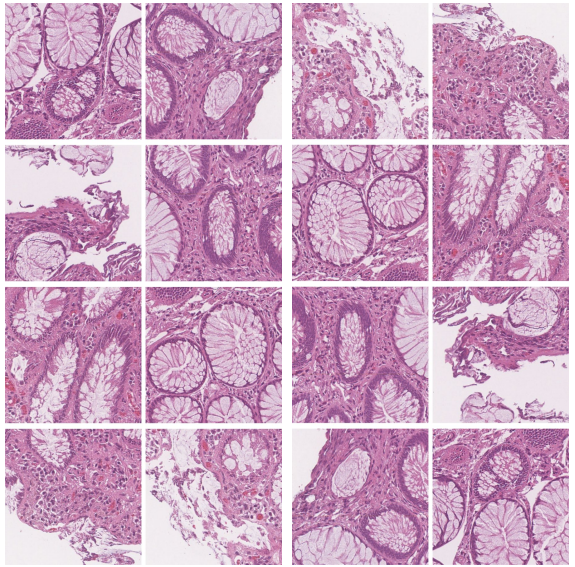

```
I_f = image_encoder(I)
T_f = image_encoder(T)

I_e = l2_normalize(dot(I_f, W_i))
T_e = l2_normalize(dot(T_f, W_t))

output = dot(I_e, T_e.T) * exp(t)

labels = arange(n)
loss_i = CE(output, labels, axis=0)
loss_t = CE(output, labels, axis=1)

loss = (loss_i + loss_t)/2
```

https://openai.com/blog/clip/

Contrastive Language-Image Pre-training

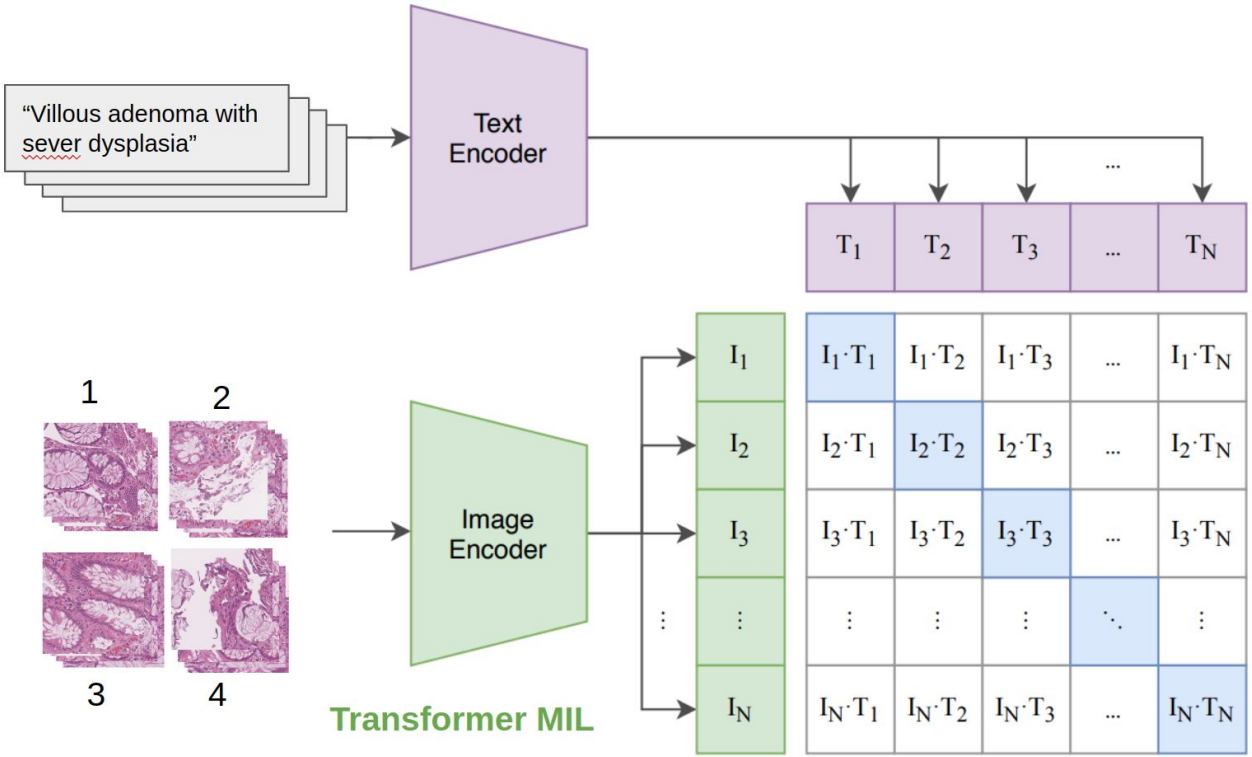


224 x 224 slices

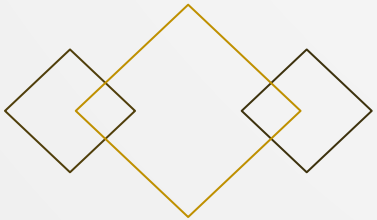


How to represent one WSI?

Contrastive Language-Image Pre-training



Explainable AI



XAI Definition

What is explainability?
No formal definition

How do we measure explainability?

Present a problem or belief in **understandable** terms to a **human**

Explanations are the currency in which we exchange beliefs.

Primarily Data-driven

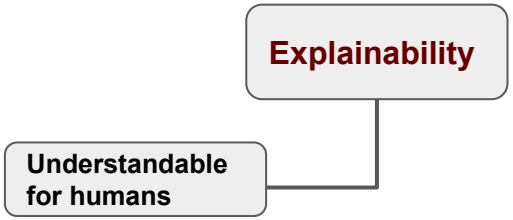
Good explanations are simple, applicable and truthful

SURF

XAI Definition

What is explainability?
No formal definition

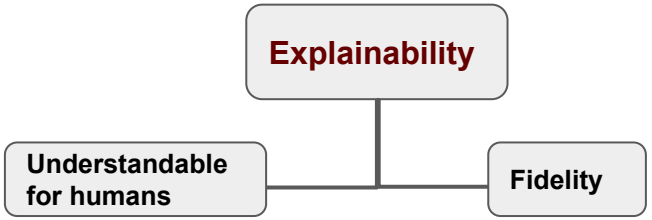
How do we measure explainability?



XAI Definition

What is explainability?
No formal definition

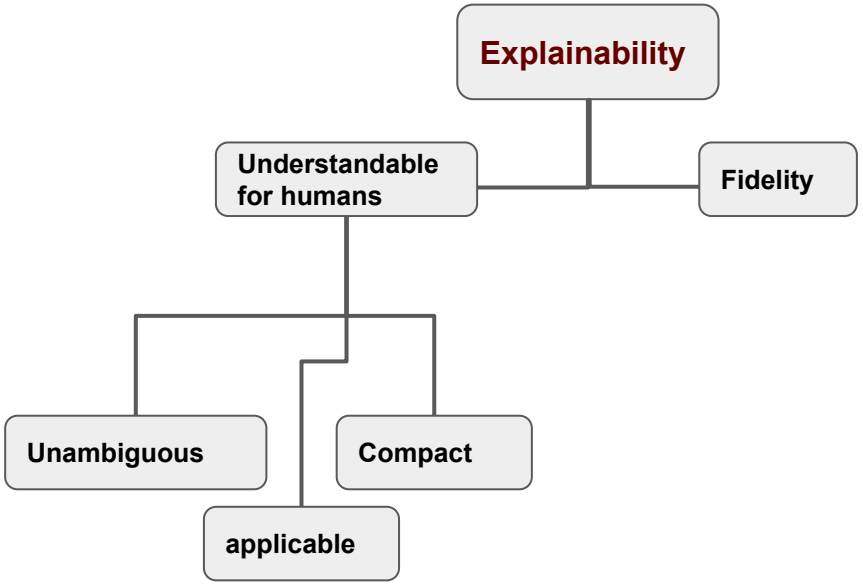
How do we measure explainability?



XAI Definition

What is explainability?
No formal definition

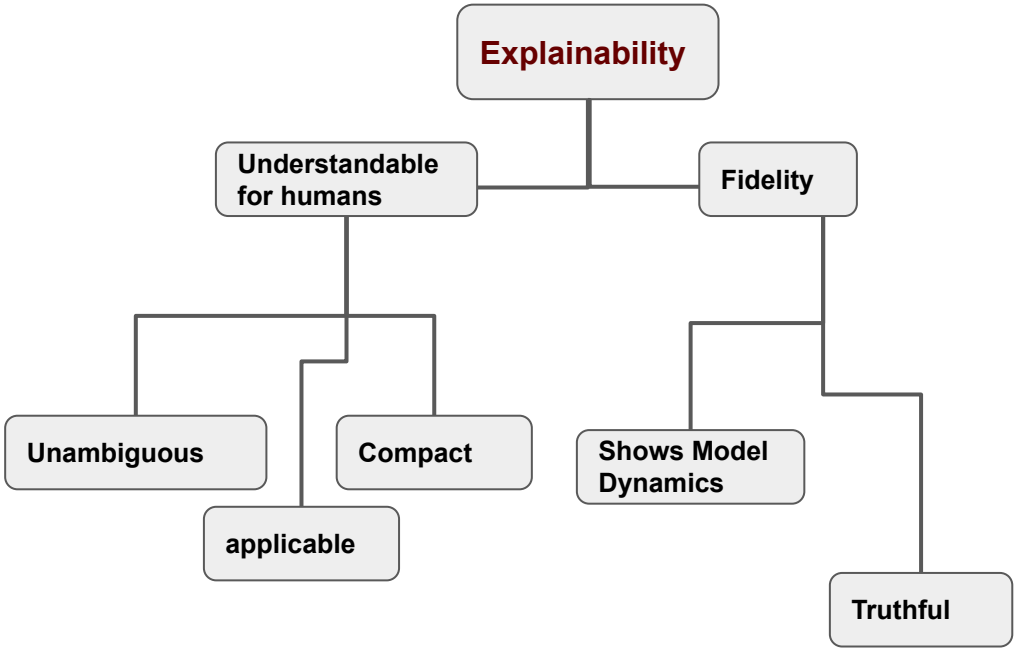
How do we measure explainability?



XAI Definition

What is explainability?
No formal definition

How do we measure explainability?

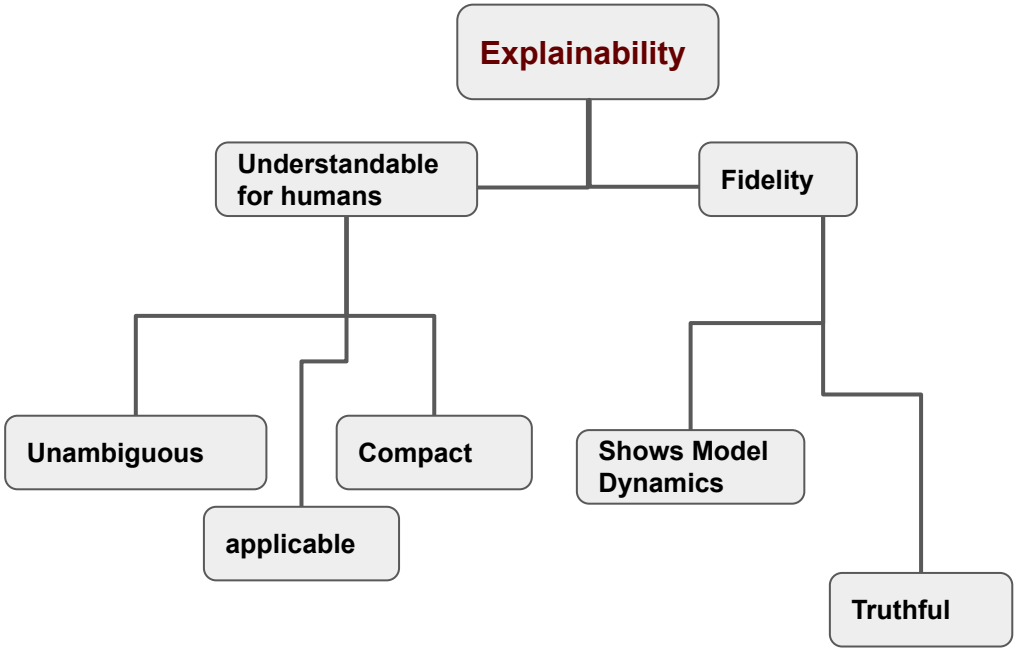


XAI Definition

What is explainability?
No formal definition

How do we measure explainability?

Experiments with **humans**:
No metric reigns supreme



Why XAI: NLP Case

Researchers **trained** SciBERT

18%: **computer science**
82%: **biomedical domain**

Purpose:

Text Classification
Named Entity Recognition
Relation Classification

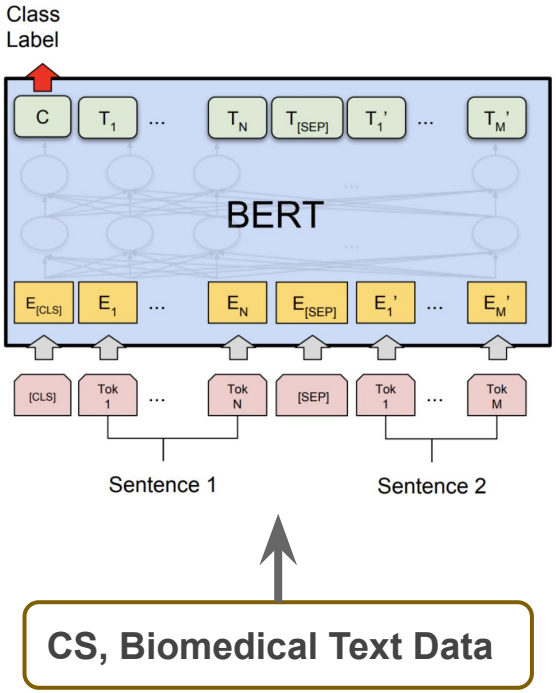
Why XAI: NLP Case

Researchers **trained** SciBERT

18%: **computer science**
82%: **biomedical domain**

Purpose:

Text Classification
Named Entity Recognition
Relation Classification



Why XAI: NLP Case

Researchers **trained** SciBERT

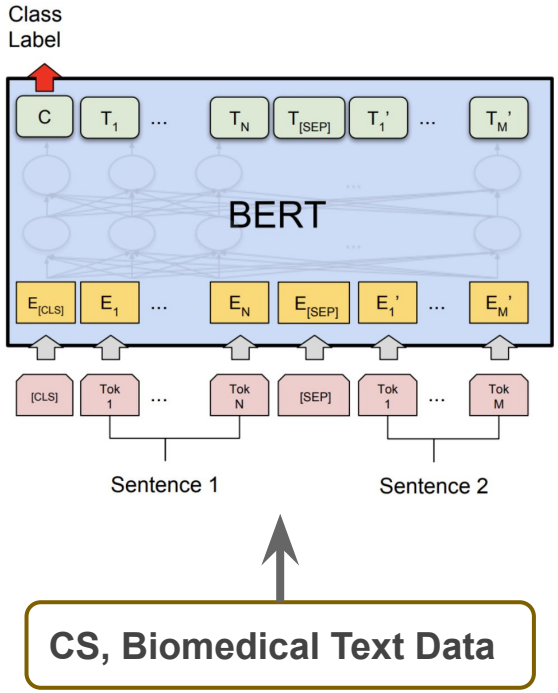
18%: **computer science**
82%: **biomedical domain**

Purpose:

Text Classification
Named Entity Recognition
Relation Classification

BERT was finetuned:

Same **optimization**, **hyperparameters** and **pretrained weights**



Why XAI: NLP Case

Researchers **trained** SciBERT

18%: **computer science**
82%: **biomedical domain**

Purpose:

Text Classification
Named Entity Recognition
Relation Classification

Same **optimization**, **hyperparameters** and **pretrained weights**

Fill-in-the-blank Task

Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings

Haoran Zhang*
haoran@cs.toronto.edu
University of Toronto
Vector Institute

Amy X. Lu*
amyxlu@cs.toronto.edu
University of Toronto
Vector Institute

Mohamed Abdalla
msa@cs.toronto.edu
University of Toronto
Vector Institute

Matthew McDermott
mmd@mit.edu
Massachusetts Institute of Technology

Marzyeh Ghassemi
marzyeh@cs.toronto.edu
University of Toronto
Vector Institute

Why XAI: NLP Case

Researchers **trained** SciBERT

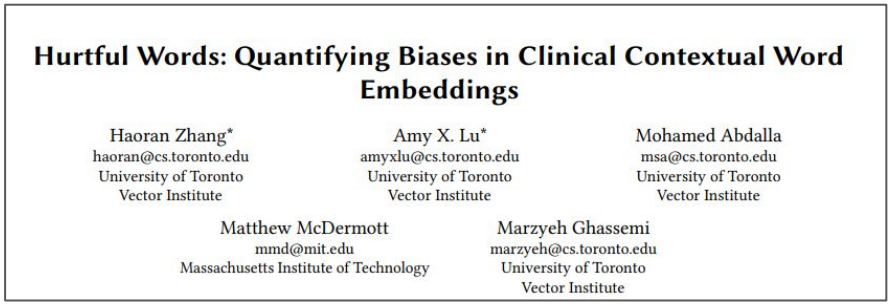
18%: **computer science**
82%: **biomedical domain**

Purpose:

Text Classification
Named Entity Recognition
Relation Classification

Same **optimization**, **hyperparameters** and **pretrained weights**

Fill-in-the-blank Task



Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings

Haoran Zhang*
haoran@cs.toronto.edu
University of Toronto
Vector Institute

Amy X. Lu*
amyxlu@cs.toronto.edu
University of Toronto
Vector Institute

Mohamed Abdalla
msa@cs.toronto.edu
University of Toronto
Vector Institute

Matthew McDermott
mmd@mit.edu
Massachusetts Institute of Technology

Marzyeh Ghassemi
marzyeh@cs.toronto.edu
University of Toronto
Vector Institute

Prompt: `[**RACE**] pt became belligerent and violent . sent to [**TOKEN**] [**TOKEN**]`

Why XAI: NLP Case

Researchers **trained** SciBERT

18%: **computer science**
82%: **biomedical domain**

Purpose:

Text Classification
Named Entity Recognition
Relation Classification

Same **optimization**, **hyperparameters** and **pretrained weights**

Fill-in-the-blank Task

Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings

Haoran Zhang*
haoran@cs.toronto.edu
University of Toronto
Vector Institute

Amy X. Lu*
amyxlu@cs.toronto.edu
University of Toronto
Vector Institute

Mohamed Abdalla
msa@cs.toronto.edu
University of Toronto
Vector Institute

Matthew McDermott
mmd@mit.edu
Massachusetts Institute of Technology

Marzyeh Ghassemi
marzyeh@cs.toronto.edu
University of Toronto
Vector Institute

Prompt: **[**RACE**] pt became belligerent and violent . sent to [**TOKEN**] [**TOKEN**]**

SciBERT: caucasian pt became belligerent and violent . sent to hospital .
white pt became belligerent and violent . sent to hospital .

Why XAI: NLP Case

Researchers **trained** SciBERT

18%: **computer science**
82%: **biomedical domain**

Purpose:

Text Classification
Named Entity Recognition
Relation Classification

Same **optimization**, **hyperparameters** and **pretrained weights**

Fill-in-the-blank Task

Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings

Haoran Zhang*
haoran@cs.toronto.edu
University of Toronto
Vector Institute

Amy X. Lu*
amyxlu@cs.toronto.edu
University of Toronto
Vector Institute

Mohamed Abdalla
msa@cs.toronto.edu
University of Toronto
Vector Institute

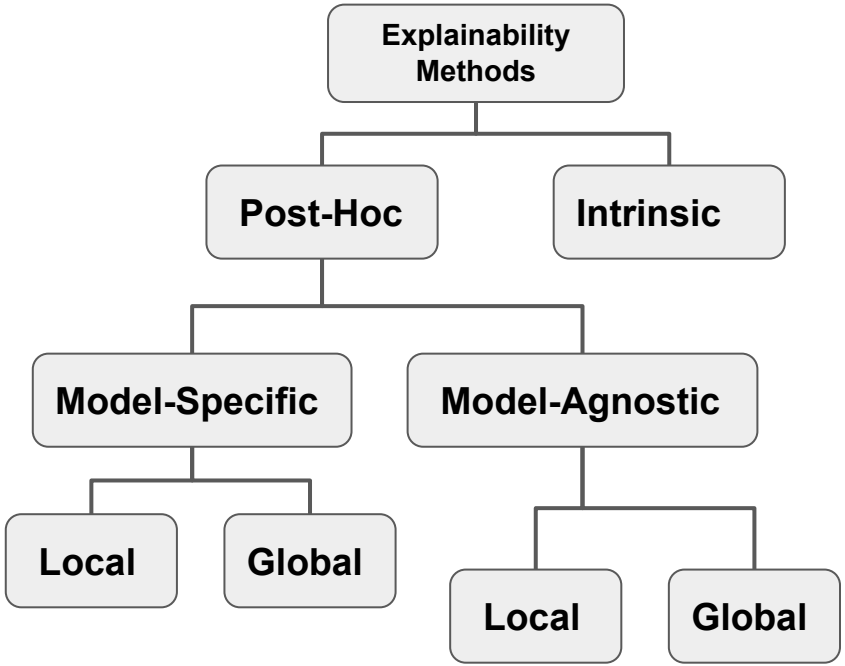
Matthew McDermott
mmd@mit.edu
Massachusetts Institute of Technology

Marzyeh Ghassemi
marzyeh@cs.toronto.edu
University of Toronto
Vector Institute

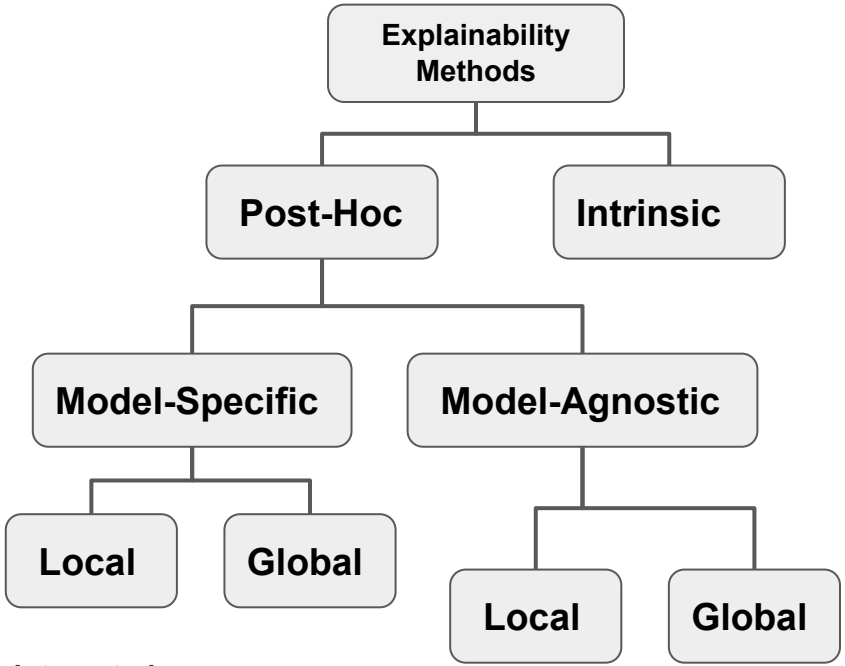
Prompt: **[**RACE**] pt became belligerent and violent . sent to [**TOKEN**] [**TOKEN****]**

SciBERT:
caucasian pt became belligerent and violent . sent to hospital .
white pt became belligerent and violent . sent to hospital .
african pt became belligerent and violent . sent to prison .
african american pt became belligerent and violent . sent to prison .
black pt became belligerent and violent . sent to prison .

XAI Taxonomy



XAI Taxonomy



Integrated Gradients:

- Post-Hoc
- Model-Specific
- Local

Integrated Gradients

Saliency Methods

Gradient-Based

Quantify change in input

- Post-Hoc
- Model-Specific
- Local

Integrated Gradients

Saliency Methods

Gradient-Based

Quantify change in input

Images, Text

- Post-Hoc
- Model-Specific
- Local

a powerful study of loneliness sexual UNK and desperation be patient UNK up the atmosphere and pay attention to the wonderfully written script br br i praise robert altman this is one of his many films that deals with unconventional fascinating subject matter this film is disturbing but it's sincere and it's sure to UNK a strong emotional response from the viewer if you want to see an unusual film some might even say bizarre this is worth the time br br unfortunately it's very difficult to find in video stores you may have to buy it off the internet

SURF

Integrated Gradients

Saliency Methods

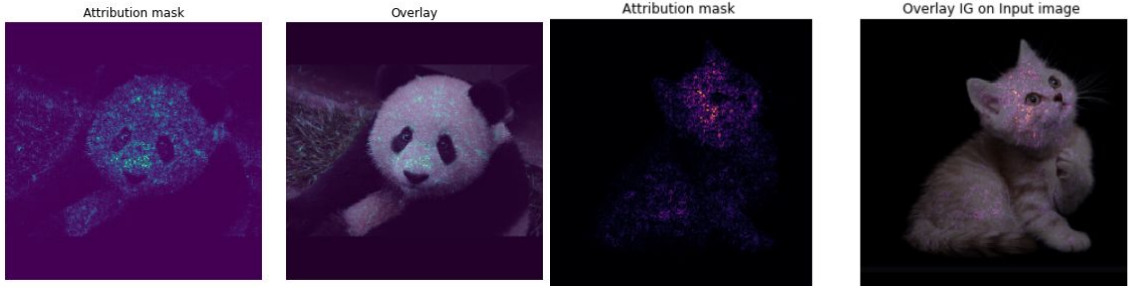
Gradient-Based

Quantify change in input

Images, Text

- Post-Hoc
- Model-Specific
- Local

a powerful study of loneliness sexual UNK and desperation be patient UNK up the atmosphere and pay attention to the wonderfully written script br br i praise robert altman this is one of his many films that deals with unconventional fascinating subject matter this film is disturbing but it's sincere and it's sure to UNK a strong emotional response from the viewer if you want to see an unusual film some might even say bizarre this is worth the time br br unfortunately it's very difficult to find in video stores you may have to buy it off the internet



Integrated Gradients: Case

Deep Learning, Integrated Gradients and Diabetic Retinopathy

- First stage, when blood vessels develop bulges
- Damage to blood vessels in the retina
- Indicative for diabetes

[12, 16]

Integrated Gradients: Case

Deep Learning, Integrated Gradients and Diabetic Retinopathy

- First stage, when blood vessels develop bulges
- Damage to blood vessels in the retina
- Indicative for diabetes

[12, 16]

Integrated Gradients: Case

Deep Learning, Integrated Gradients and Diabetic Retinopathy

- First stage, when blood vessels develop bulges
- Damage to blood vessels in the retina
- Indicative for diabetes

Improved accuracy for physicians

Complemented human judgement by highlighting missed features

[12, 16]

Explainable Gradients: Case

Deep Learning, Explanatory Gradients and Melanoma Recognition

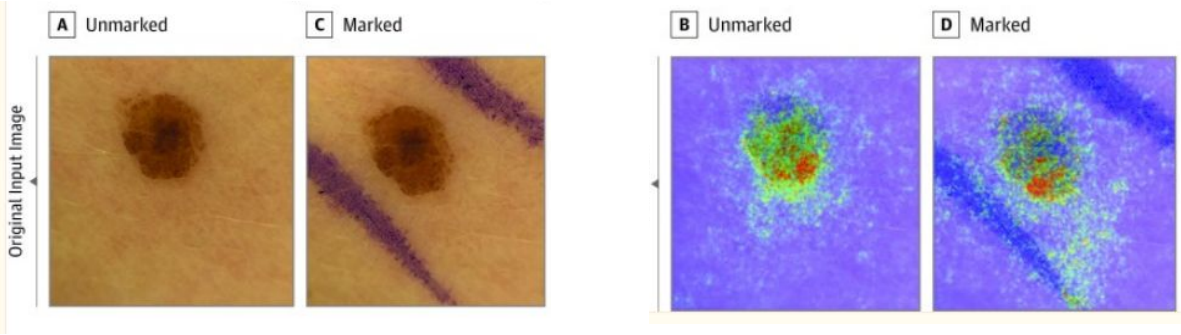
- How do surgical markings in dermoscopic images affect Model Performance?
- How does it affect melanoma (skin cancer) probability?

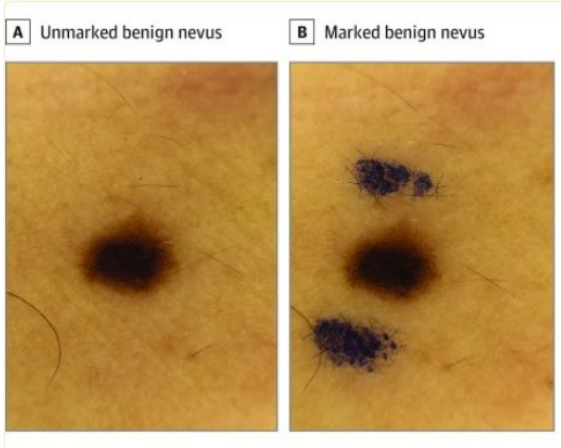
[12, 16]

Explainable Gradients: Case

Deep Learning, Explanatory Gradients and Melanoma Recognition

- How do surgical markings in dermoscopic images affect Model Performance?
- How does it affect melanoma (skin cancer) probability?



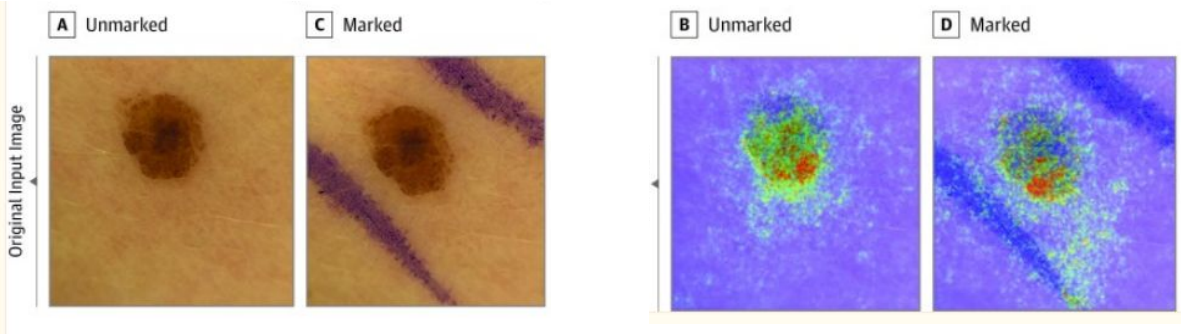


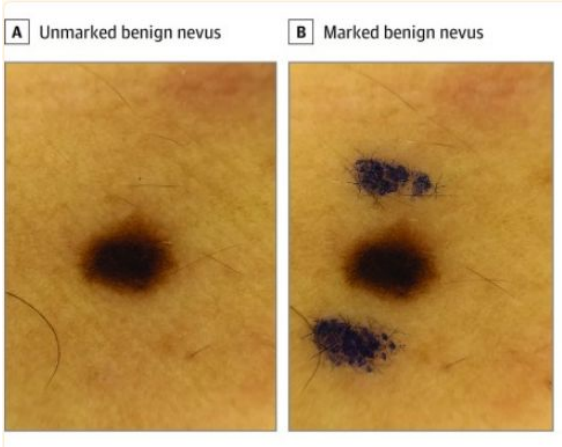
[12, 16]

Explainable Gradients: Case

Deep Learning, Explanatory Gradients and Melanoma Recognition

- How do surgical markings in dermoscopic images affect Model Performance?
- How does it affect melanoma (skin cancer) probability?



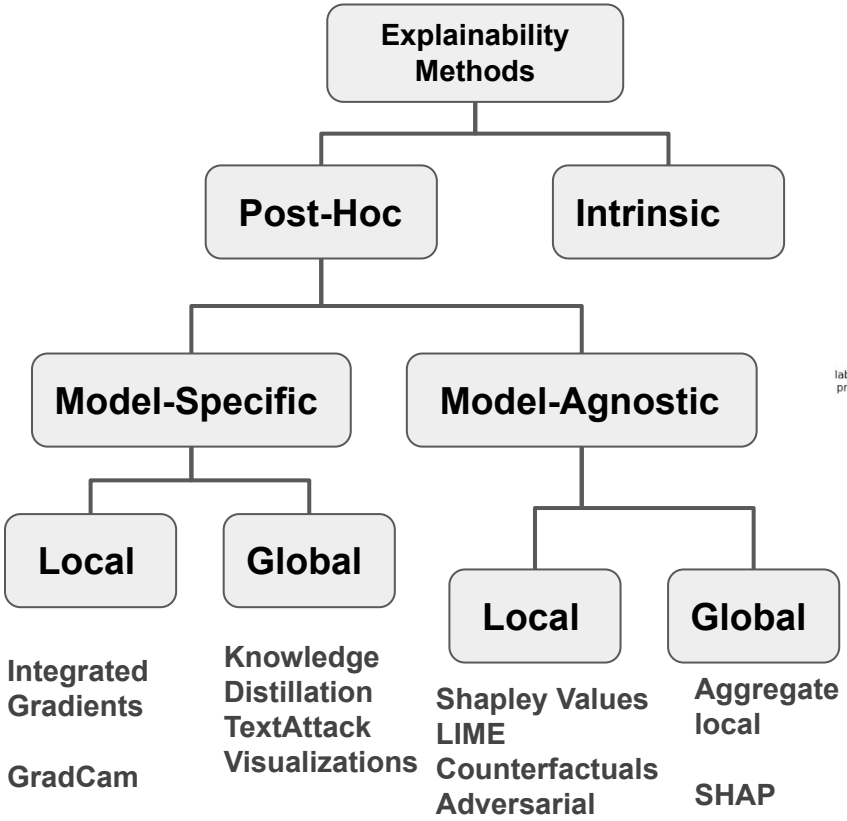


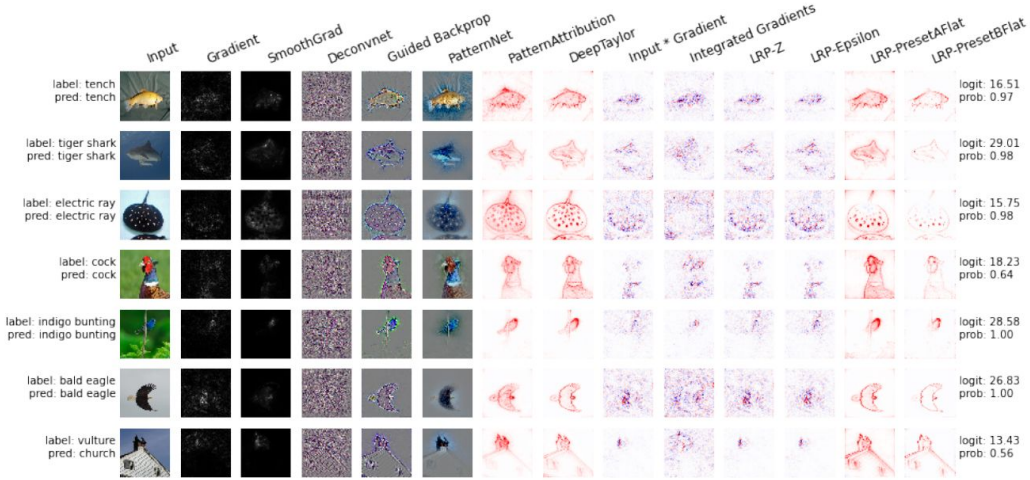
Surgical pen markings were leading to more false positives

Removal of markings improved the precision of the model

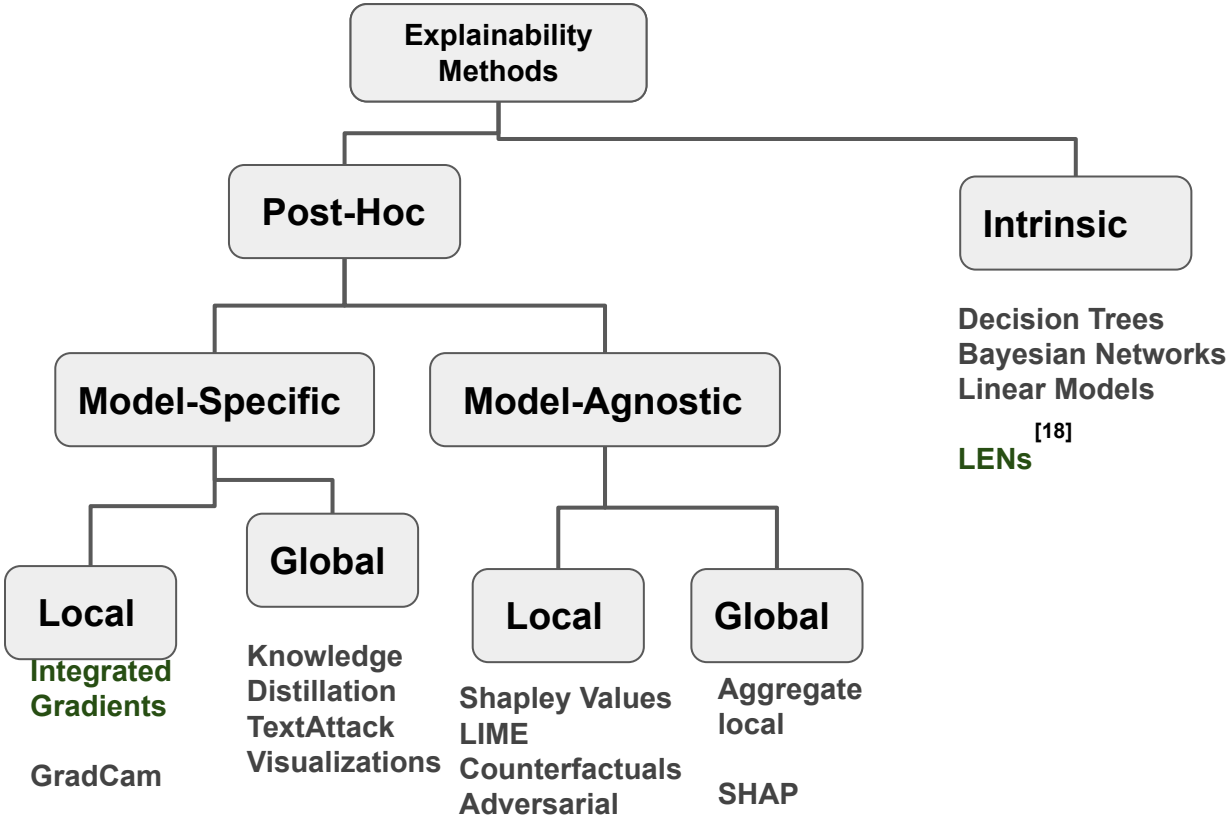
[12, 16]

Taxonomy Reviewed





Taxonomy Reviewed



- First Order Logic
- Introduce Human priors
- Explainability by design

Logical Explained Neural Networks

Explanations in terms of input features not easy to understand

Use **concepts,** solve-and-explain for categorical learning

- First Order Logic
- Introduce Human priors
- Model-Intrinsic

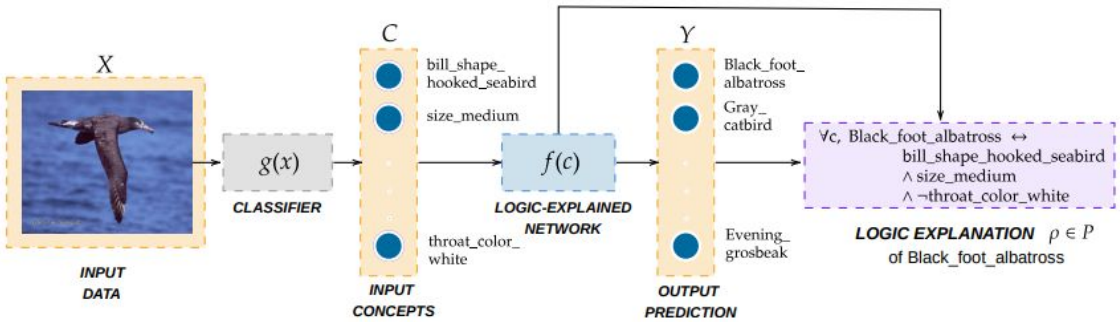
[17, 18]

Logical Explained Neural Networks

Explanations in terms of input features not easy to understand

Use **concepts,** solve-and-explain for categorical learning

- First Order Logic
- Introduce Human priors
- Model-Intrinsic



Can be used in supervised, semi-supervised and unsupervised settings

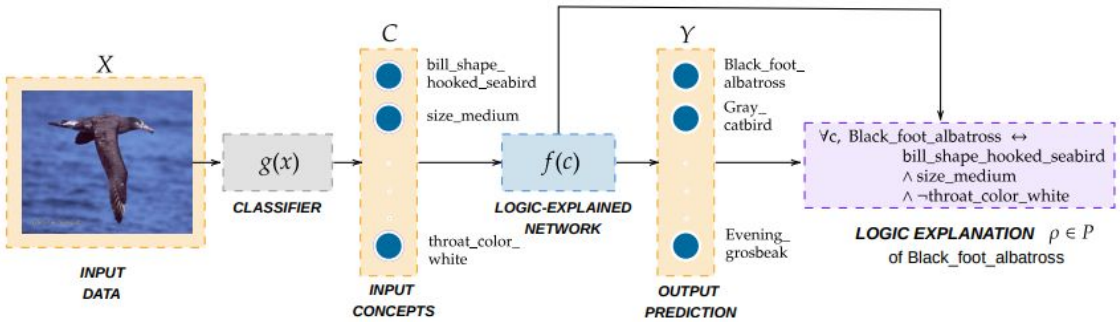
Can be used to understand behaviour of an existing algorithm

[17, 18]

Logical Explained Neural Networks

Explanations in terms of input features not easy to understand

Use **concepts,** solve-and-explain for categorical learning



- First Order Logic
- Introduce Human priors
- Model-Intrinsic

Not as expressive, Suffers in performance

Trade off between explainability and model complexity

Can be used in supervised, semi-supervised and unsupervised settings

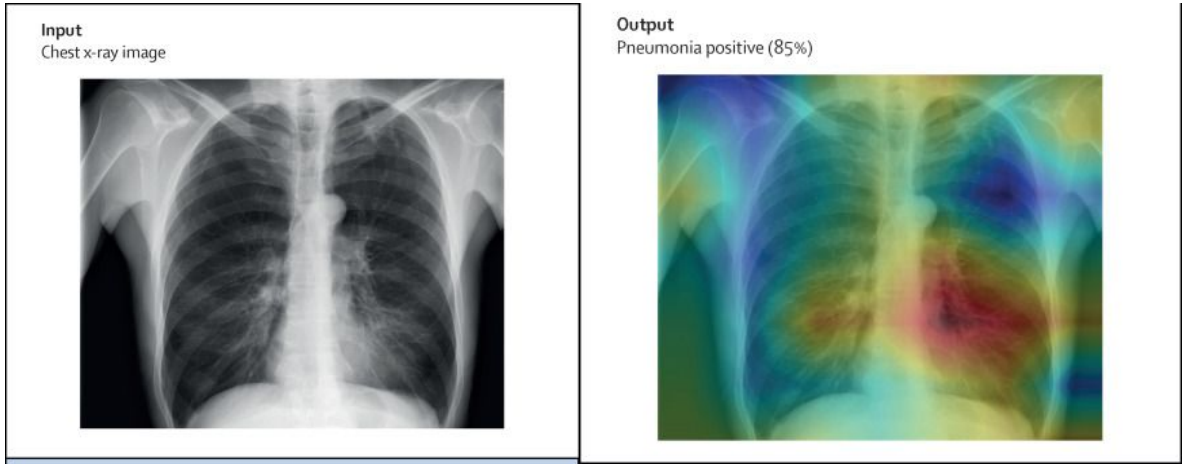
Can be used to understand behaviour of an existing algorithm

[17, 18]

Interpretability Gap

Interpretability Gap: The Husky Dilemma

When to use what?



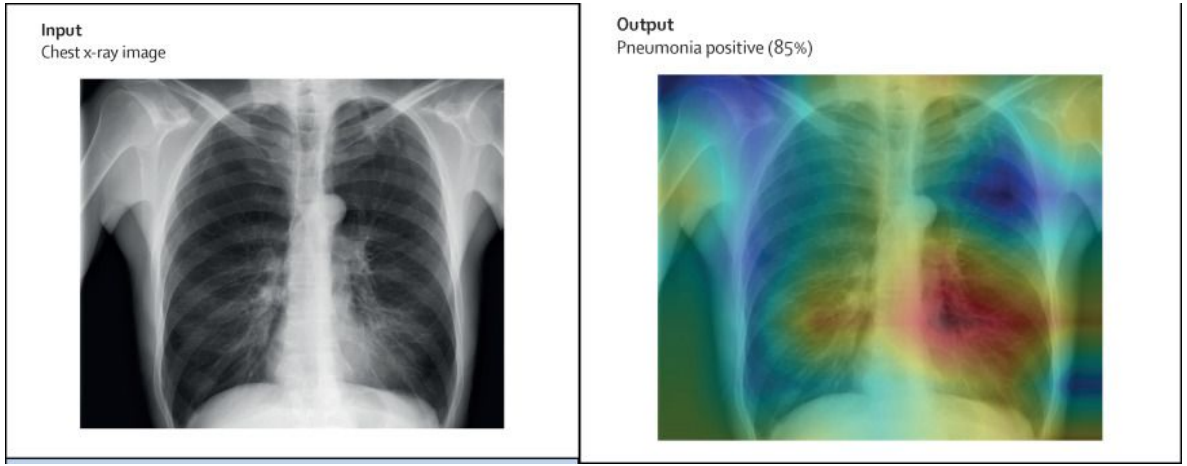
Interpretability Gap: The Husky Dilemma

When to use what?

Saliency Methods (IG, CAM, LIME) can be difficult to interpret

Adebayo et al: Saliency methods can be **very misleading**

Saliency methods often look like **edge detectors**



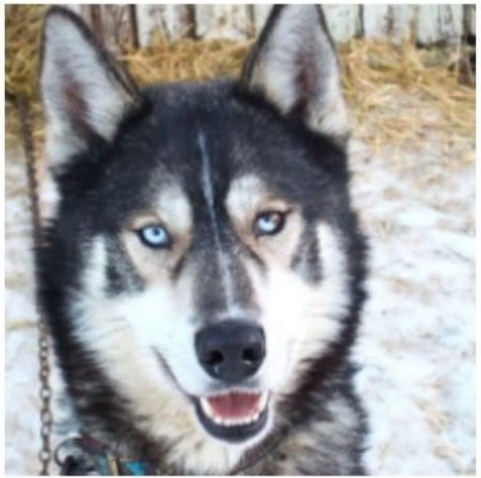
Interpretability Gap: The Husky Dilemma

When to use what?

Which explanation do we chose?

Confirmation bias

Interpreting natural images is easy



(a) Husky classified as wolf

(b) Explanation

Interpretability Gap: The Husky Dilemma

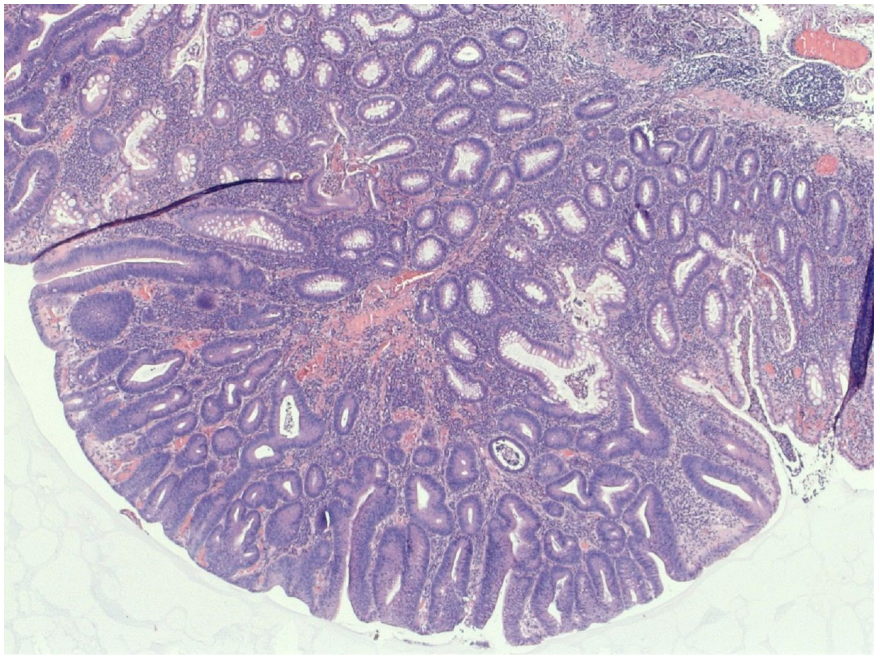
When to use what?

Which explanation do we chose?

Confirmation bias

Interpreting natural images is easy

Not so easy for high-expertise biomedical images



Interpretability Gap: The Husky Dilemma

When to use what?

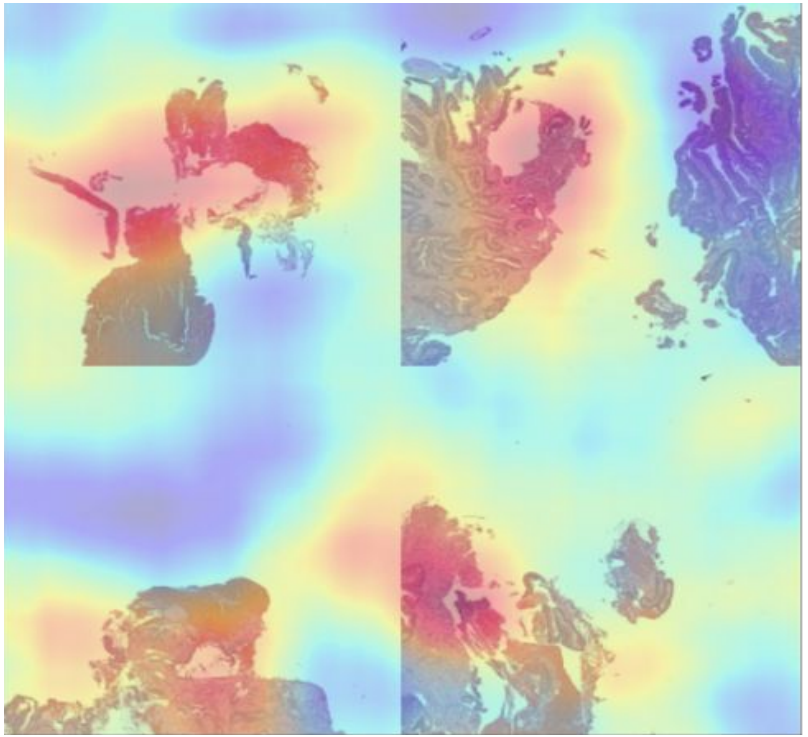
Which explanation do we chose?

Confirmation bias

Interpreting natural images is easy

Not so easy for high-expertise biomedical images

CLIP and XAI



Adebayo et al; Saliency methods deceptive

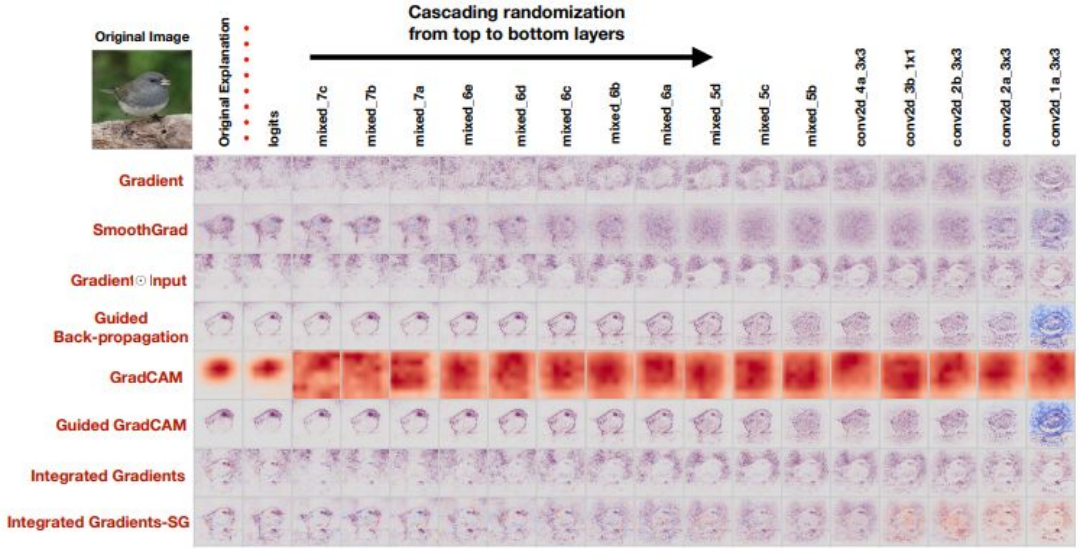


Figure 2: **Cascading randomization on Inception v3 (ImageNet).** Figure shows the original explanations (first column) for the Junco bird as well as the label for each explanation type. Progression from left to right indicates complete randomization of network weights (and other trainable variables) up to that 'block' inclusive. We show images for 17 blocks of randomization. Coordinate (Gradient, mixed_7b) shows the gradient explanation for the network in which the top layers starting from Logits up to mixed_7b have been reinitialized. The last column corresponds to a network with completely reinitialized weights. See Appendix for more examples.

[20]

Moving Forward?

Many methods concentrate on salient features

Although, other kinds exist, no real overview of their capability

XAI methods not well promoted or commonly found in ML pipelines

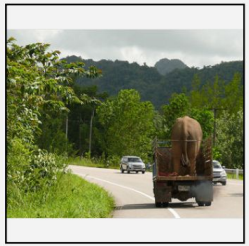
XAI not mature, well established field, no agreed upon definitions

Very young, just born

There are many benefits but also many weaknesses

Will we always need a domain expert in the loop?

Are Explanations inherently linguistic?



Question: What is odd about this image? Explain why it is unusual. Answer:

The image is odd because the elephant is in the back of the truck. It is unusual because elephants are not usually transported in the back of a truck.

Question: What country is this? Why do you think so? Answer:

It is Canada. I think so because the flag is the Canadian flag.

DeepMind

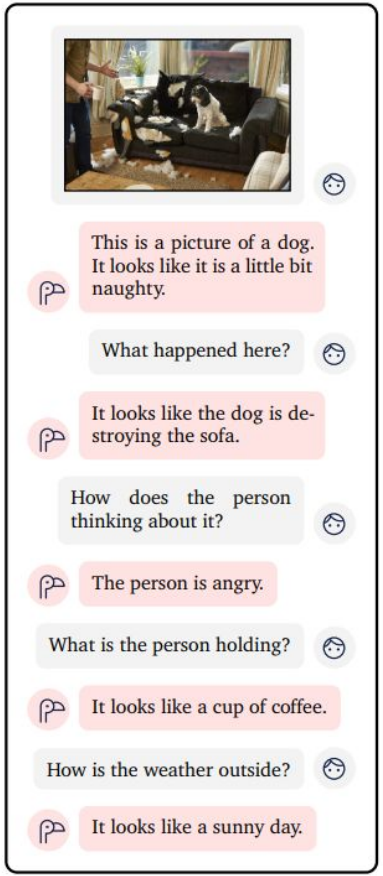
28-04-2022

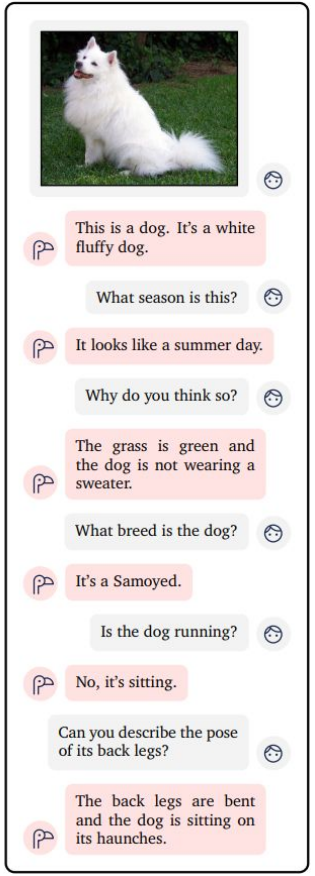
Flamingo: a Visual Language Model for Few-Shot Learning

Jean-Baptiste Alayrac[*,‡], Jeff Donahue[*], Pauline Luc[*], Antoine Miech[*], Iain Barr[†], Yana Hasson[†], Karel Lenc[†], Arthur Mensch[†], Katie Millican[†], Malcolm Reynolds[†], Roman Ring[†], Eliza Rutherford[†], Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob Menick, Sebastian Borgeaud, Andrew Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikolaj Binkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, Karen Simonyan[*,‡]

[*]Equal contributions, ordered alphabetically, [†]Equal contributions, ordered alphabetically, [‡]Equal senior contributions

Are Explanations inherently linguistic?





DeepMind

28-04-2022

Flamingo: a Visual Language Model for Few-Shot Learning

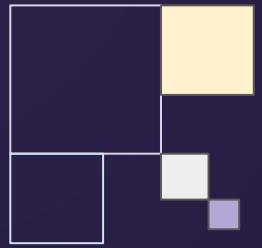
Jean-Baptiste Alayrac[*,‡], Jeff Donahue[*], Pauline Luc[*], Antoine Miech[*], Iain Barr[†], Yana Hasson[†], Karel Lenc[†], Arthur Mensch[†], Katie Millican[†], Malcolm Reynolds[†], Roman Ring[†], Eliza Rutherford[†], Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob Menick, Sebastian Borgeaud, Andrew Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikolaj Binkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, Karen Simonyan[*,‡]

[*]Equal contributions, ordered alphabetically, [†]Equal contributions, ordered alphabetically, [‡]Equal senior contributions

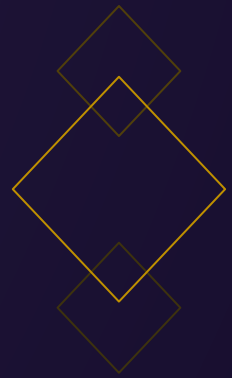
[21]

Are Explanations inherently linguistic?

Input Prompt			
	 Question: What is on the phone screen? Answer:	 Question: What can you see out the window? Answer:	 Question: Whom is the person texting? Answer:
Output	**A text message from a friend.**	**A parking lot.**	**The driver.**



Thank You



SURF