Generative Forms of Multimedia Content

Alan F. Smeaton

Dublin City University

Invited Keynote talk presented at IEEE Conference on Multimedia and Expo (ICME)

7th July 2020

A World Leading SFI Research Centre

Science Foundation Ireland For what's next

Insight

SFI RESEARCH CENTRE FOR DATA ANALYTICS

HOST INSTITUTIONS

DCU, NUI Galway OÉ Gaillimh, UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh, UCD Dublin

PARTNER INSTITUTIONS

Maynooth University National University of Ireland Maynooth, Tyndall National Institute, Trinity College Dublin Coláiste na Tríonóide, Baile Átha Cliath The University of Dublin, University of Limerick Ollscoil Luimnigh

FUNDED BY:

European Union European Regional Development Fund, Science Foundation Ireland For what's next

Generative Forms of Multimedia Content

Part 1

Alan F. Smeaton

Dublin City University

Invited Keynote talk presented at IEEE Conference on Multimedia and Expo (ICME)

7th July 2020

A World Leading SFI Research Centre

Science Foundation Ireland SFI For what's next

Insight

SFI RESEARCH CENTRE FOR DATA ANALYTICS

HOST INSTITUTIONS

DCU

NUI Galway OÉ Gaillimh

UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh

UCD DUBLIN

PARTNER INSTITUTIONS

Maynooth University National University of Ireland Maynooth

Tyndall National Institute

Trinity College Dublin Coláiste na Tríonóide, Baile Átha Cliath The University of Dublin

UNIVERSITY OF LIMERICK OLLSCOIL LUIMNIGH

FUNDED BY:

European Union European Regional Development Fund

Science Foundation Ireland SFI For what's next

Alan Smeaton ?

- DCU Professor
- A Founding Director of Insight
- Worked on many research projects across domains so breadth is important
- Started in NLP for IR, then image, video, so background in multimedia analysis, indexing and search
- The route that taught me ML
- TRECVid anyone ?
- And now .. memory, memorability, helping us remember as well as plugging the gaps when we forget

Today's Talk ?

- Part 1 … AI, ML is its driver, CV and MM fostered it, issues with data bias and availability so we do data augmentation
- Part 2 … generative MM, some stunning examples then some AI art/MM tools, finishing with whether they are AI or not

- Peppered with examples, many of our own

AI Definitions

- AI definition is old, changes as tech developed
- Several dictionaries define AI as …
  (taken from Forbes Magazine)
  - "The theory and development of computer systems able to perform tasks **normally requiring human intelligence**, such as visual perception, speech recognition, decision-making, and translation between languages."
  - "A branch of computer science dealing with the simulation of **intelligent behaviour** in computers."
  - "The capability of a machine to **imitate intelligent human** behaviour."
  - "the ability of a digital computer or computer-controlled robot to perform **tasks** commonly **associated with intelligent** beings."
  - "The field of computer science dedicated to solving cognitive problems commonly **associated with human intelligence**, such as learning, problem solving, and pattern recognition."

AI Definitions

- All correct, but the fact there are so many means is an agreed topic
- What's common is that AI copies intelligent human behavior in some way
- In the early days, there was great hope for computerised intelligence .. speech, vision, language translation, etc.
- Expectations didn't materialise as the technology wasn't mature enough … so we had Winters and Springs and Summers

ML – A Brief History

- Machine Learning is one of AI's great enablers
  .. learn and replicate patterns from data
- Developed in an environment where data was structured, plentiful and "clean" .. first powerful uses were in search engines
- Clickthroughs were mined, turned into AdWords, pushed the boundaries of ML
- Through the 1990s and 2000s … ML mostly under the radar but small numbers of us were using ML for different kinds of applications

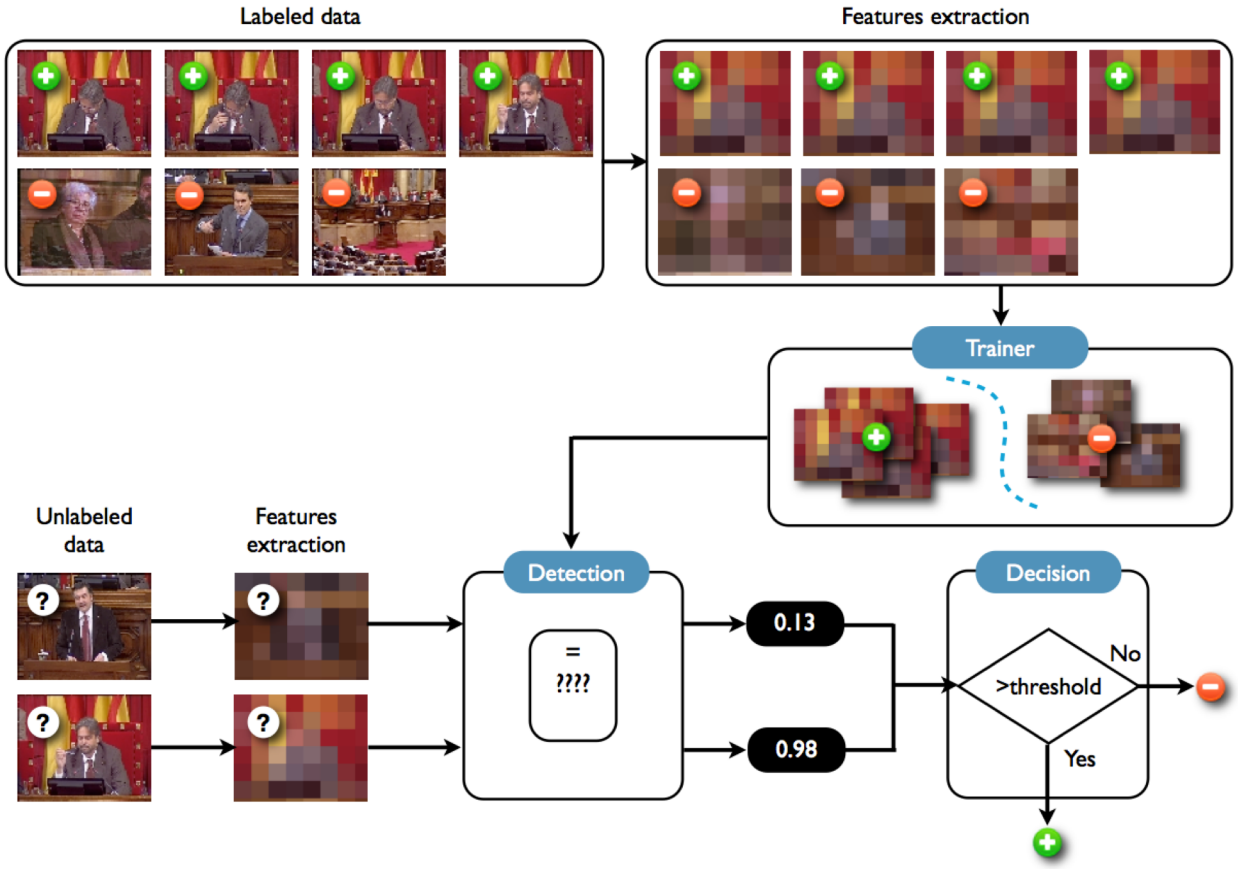
Machine Learning

- There are a host of categories of ML
- They differ in sophistication and complexity, in the amount of training data they require, in effectiveness, in compute time … they capture relationships between variables that are more subtle, more complex, than straight correlations

Machine Learning & Multimedia

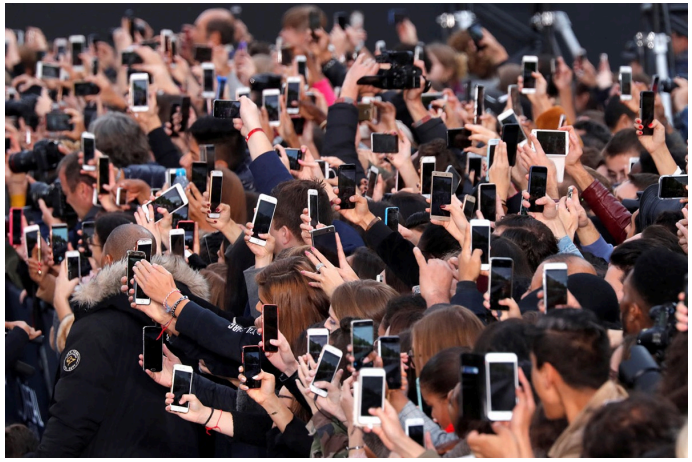
- Starting in the 1990s internet search companies used ML in-house (had the data, and computing resources) and slowly this crept into publicly-funded research
- By mid-2000s ML was used in public and private research
- We ran public benchmark tasks to push the boundaries, to develop new techniques, like ImageNet for image tags and TRECVid for video tasks .. we didn't have data or the computing resources of internet companies but we made progress
- ImageNet with 1,000 classes (tags) was still short of human accuracy but slowly improving, using those ML techniques like SVMs, random forests, decision trees, etc.

Machine Learning & Multimedia



Why ML for MM ?

Exponential increase of generated multimedia..



L'Oreal fashion show on Champs Elysees, Paris Fashion Week, 2017.

Motivation for ML in MM ?

…(or not).

Pope Francis, USA, 2015

Motivation for ML in MM ?

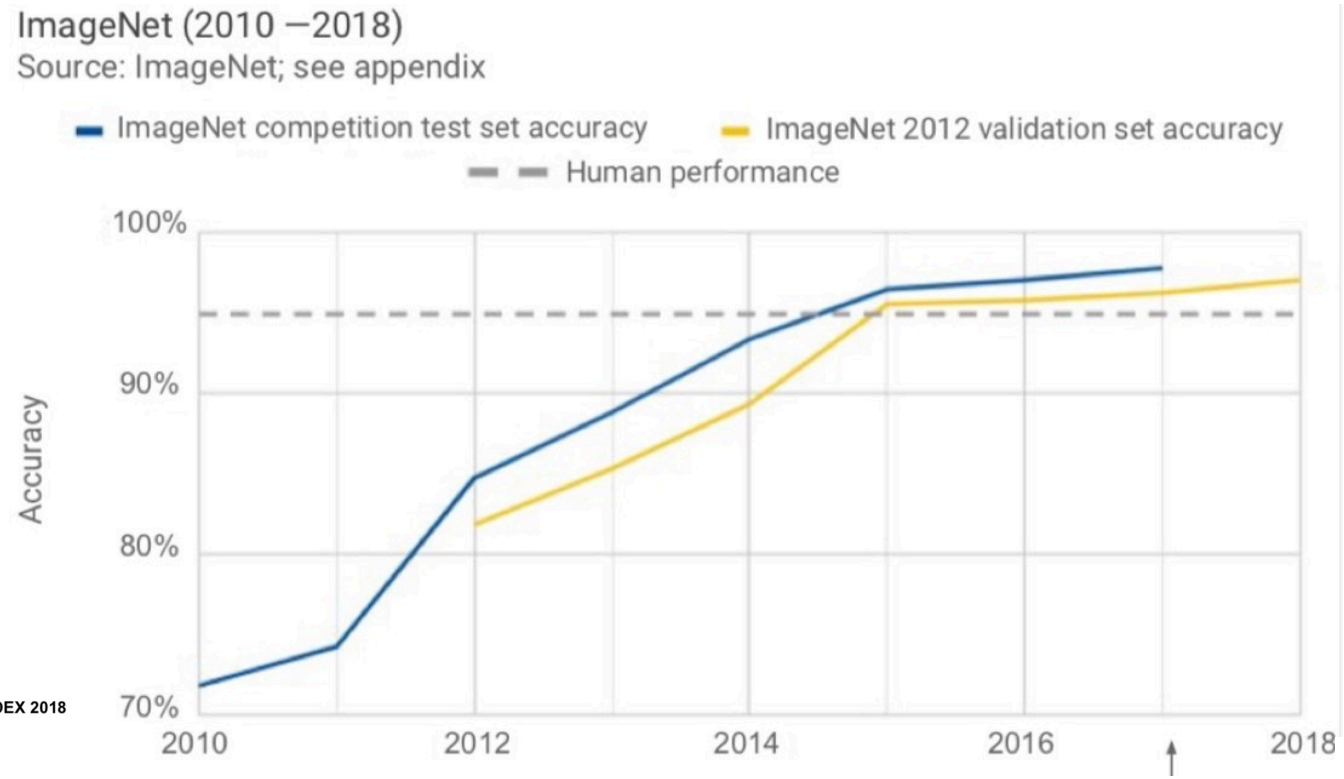
...so the challenge was to index and help retrieve this data

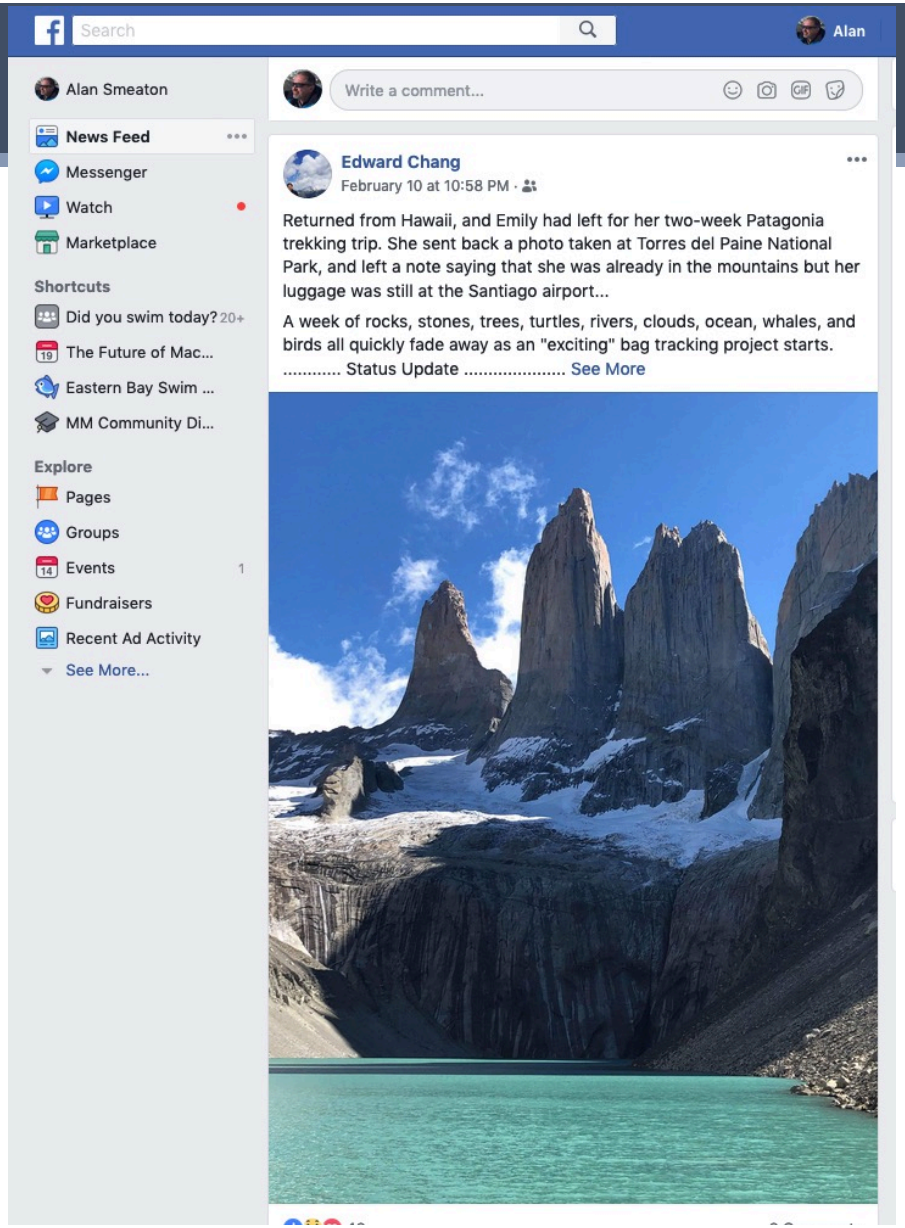
Back to ML and MM

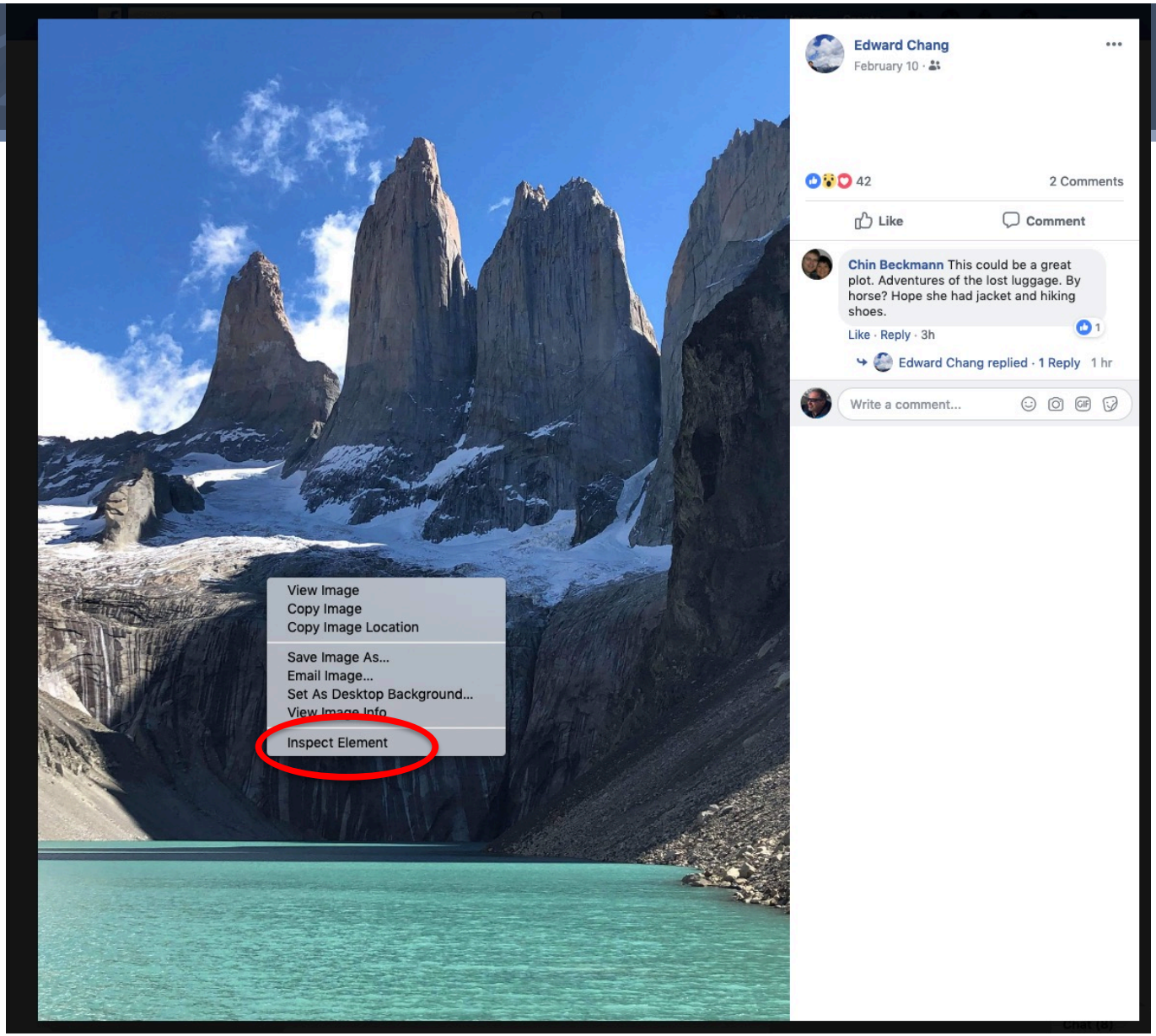
- Classification assigns a label based on features showing a discrimination between samples from each category
- CNNs are end-to-end solutions where both feature extraction and classifier training are performed at once … important later

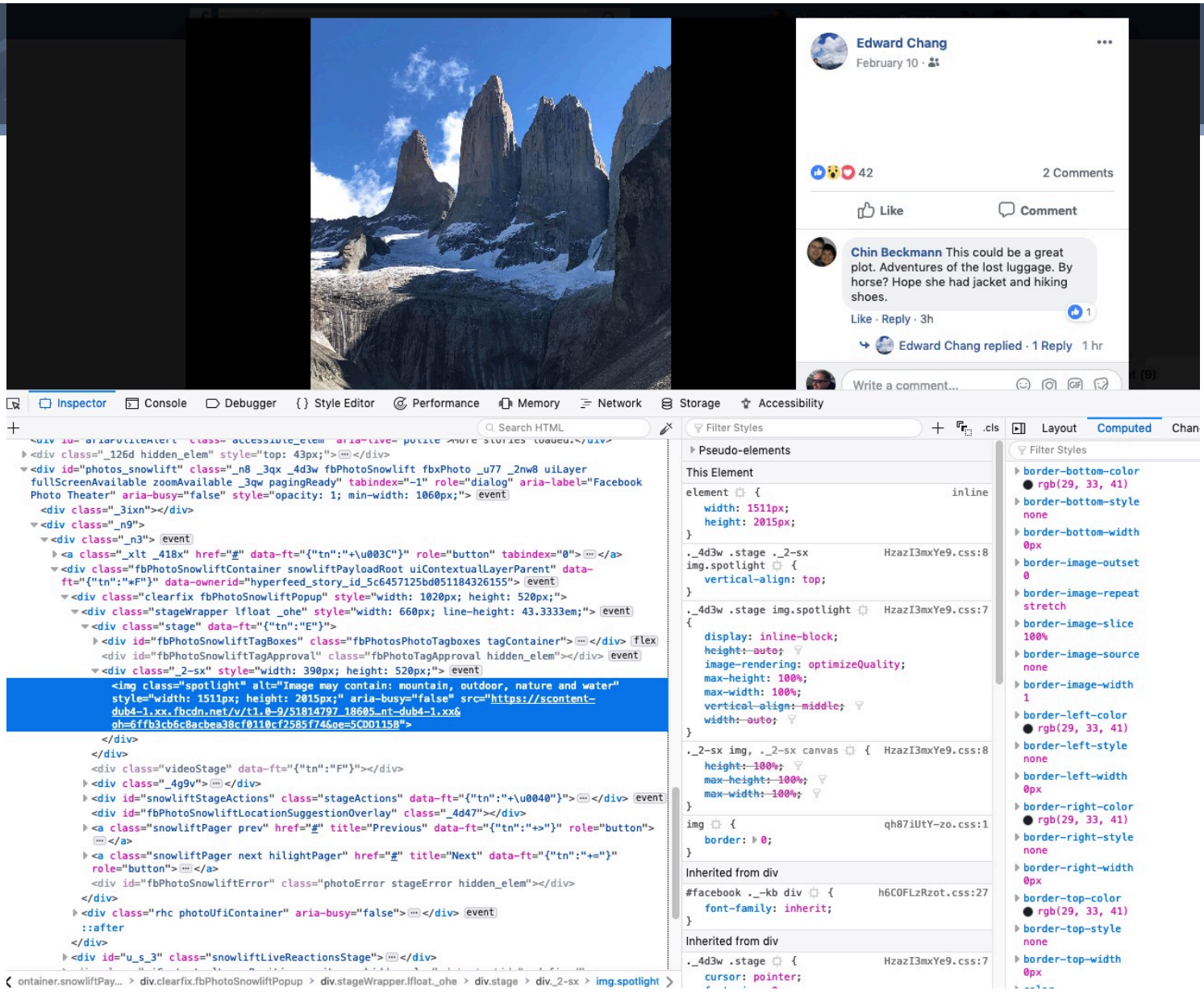
- For computer vision it works .. and really well
- **Google+** photos used it to identify objects and settings in uploaded snapshots
- **iPhone** uses it to support on-device search
- **FaceBook** uses it to tag photos

MM Indexing Improved




Search
Alan
Alan Smeaton
News Feed
Messenger
Watch
Marketplace
Shortcuts
Did you swim today? 20+
The Future of Mac...
Eastern Bay Swim ...
MM Community Di...
Explore
Pages
Groups
Events 1
Fundraisers
Recent Ad Activity
See More...
Write a comment...
Edward Chang
February 10 at 10:58 PM
Returned from Hawaii, and Emily had left for her two-week Patagonia trekking trip. She sent back a photo taken at Torres del Paine National Park, and left a note saying that she was already in the mountains but her luggage was still at the Santiago airport...
A week of rocks, stones, trees, turtles, rivers, clouds, ocean, whales, and birds all quickly fade away as an "exciting" bag tracking project starts.
..... Status Update ..... See More


View Image
Copy Image
Copy Image Location
Save Image As...
Email Image...
Set As Desktop Background...
View Image Info
Inspect Element
Edward Chang
February 10
42
2 Comments
Like
Comment
Chin Beckmann This could be a great plot. Adventures of the lost luggage. By horse? Hope she had jacket and hiking shoes.
Like · Reply · 3h
1
Edward Chang replied · 1 Reply 1 hr
Write a comment...


Edward Chang
February 10
42
2 Comments
Like
Comment
Chin Beckmann This could be a great plot. Adventures of the lost luggage. By horse? Hope she had jacket and hiking shoes.
Like · Reply · 3h
Edward Chang replied · 1 Reply 1 hr
Write a comment...
Inspector
Console
Debugger
Style Editor
Performance
Memory
Network
Storage
Accessibility
Search HTML
Filter Styles
Layout
Computed
Pseudo-elements
This Element
element { inline
width: 1511px;
height: 2015px;
img.spotlight {
vertical-align: top;
display: inline-block;
image-rendering: optimizeQuality;
max-height: 100%;
max-width: 100%;
Inherited from div
Inherited from div
cursor: pointer;

```
<img class="spotlight" alt="Image may contain: mountain, outdoor, nature and water"
style="width: 1511px; height: 2015px;" aria-bu
dub4-1.xx.fbcdn.net/v/t1.0-9/51814797_18605_nt-dub4-1.xx&
oh=6ffb3cb6c8acbea38cf0110cf2585f74&oe=5CDD115B">
```

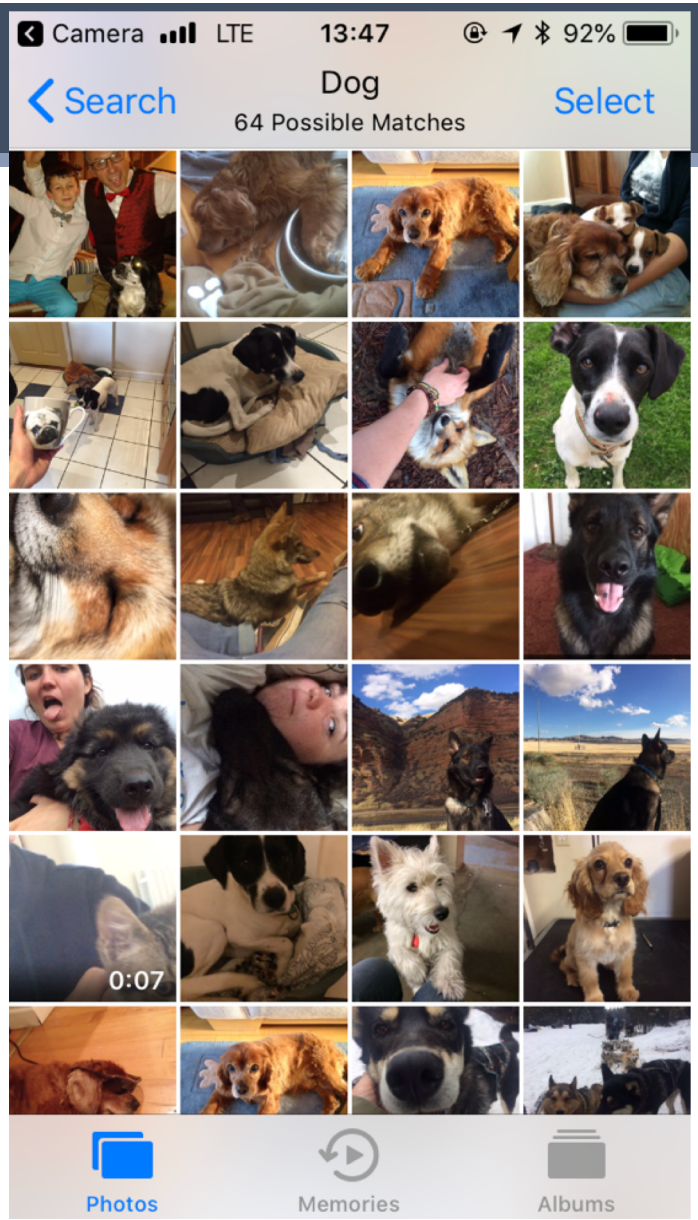
Facebook image tagging

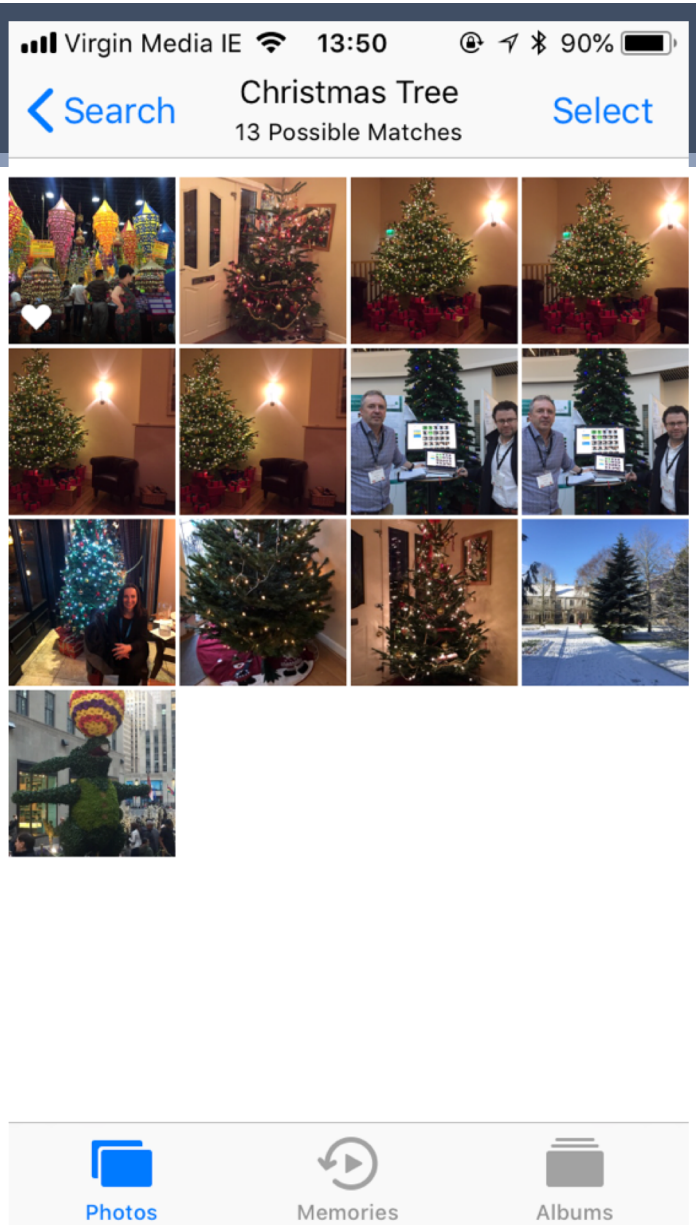
- 2 people, people smiling, indoor

- 2 people, people smiling, beard and closeup

- 1 person, sitting, and dog

15 people, people smiling, people sitting, table, food and indoor


Camera LTE 13:47 92%
Search
Dog
64 Possible Matches
Select
0:07
Photos
Memories
Albums

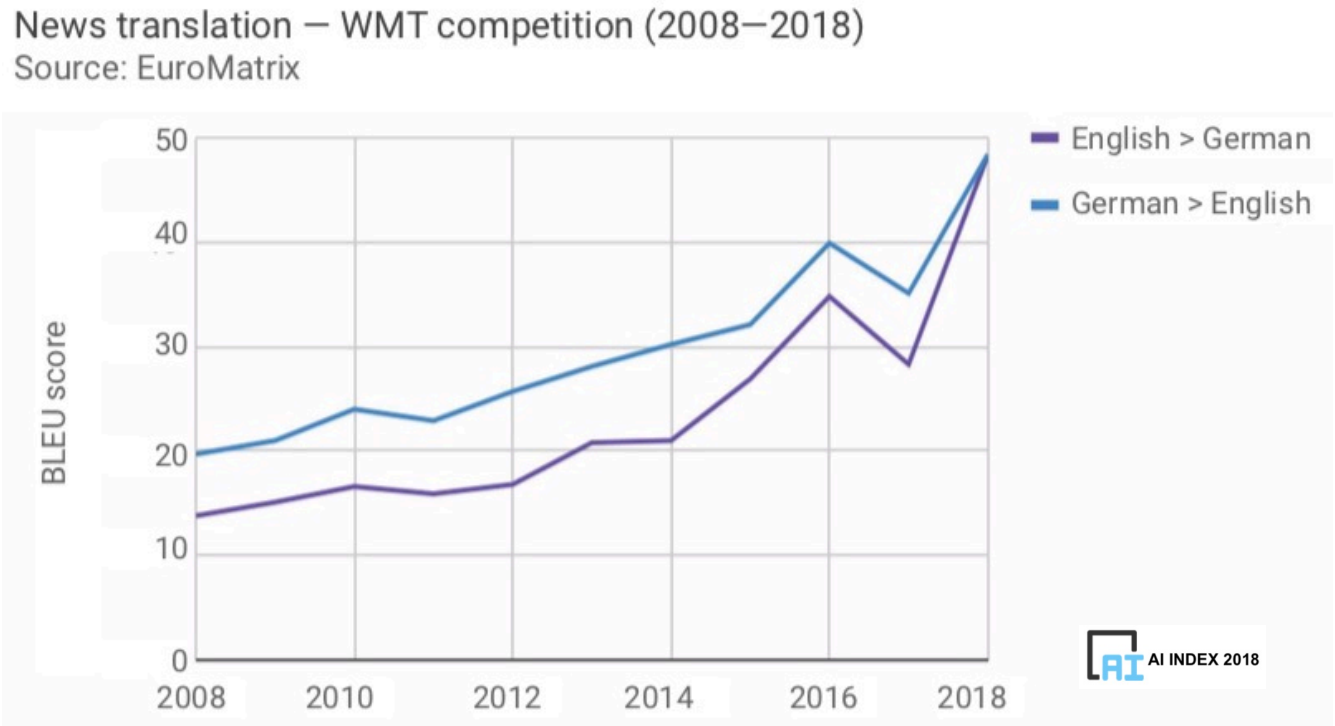

Virgin Media IE
13:50
90%
Search
Christmas Tree
13 Possible Matches
Select
Photos
Memories
Albums

AI really works – sometimes !

- We … us … MM research … half-dozen years of extreme progress

- Trickled into other areas and everybody started doing deep learning … for every known problem, so lots of bandwaggoning .. e.g. MT became statistical MT
- A zoo of ML techniques so using the most appropriate of those is a black art, a form of alchemy

Machine translation improved



Nov 2014, NY Times (premature)

Researchers Announce Advance in Image-Recognition Software

By JOHN MARKOFF NOV. 17, 2014

MOUNTAIN VIEW, Calif. — Two groups of scientists, working independently, have created artificial intelligence software capable of recognizing and describing the content of photographs and videos with far greater accuracy than ever before, sometimes even mimicking human levels of understanding.

Until now, so-called computer vision has largely been limited to recognizing individual objects. The new software, described on Monday by researchers at Google and at Stanford University, teaches itself to identify entire scenes: a group of young men playing Frisbee, for example, or a herd of elephants marching on a grassy plain.

The software then writes a caption in English describing the picture. Compared with human observations, the researchers found, the computer-written descriptions are surprisingly accurate.

The advances may make it possible to better catalog and search for the billions of images and hours of video available online, which are often poorly described and archived. At the moment, search engines like Google rely largely on written language accompanying an image or video to ascertain what it contains.

RELATED COVERAGE

Computer Eyesight Accurate AUG. 18, 20

DCU

Captioned by Human and by Google's Experimental Program

Human: "A group of men playing Frisbee in the park."
Computer model: "A group of young people playing a game of Frisbee."

1 of 6

Captioned by Human and by Google's Experimental Program

Human: "A young hockey player playing in the ice rink."
Computer model: "Two hockey players are fighting over the puck."

2 of 6

Captioned by Human and by Google's Experimental Program

Human: "A green monster kite soaring in a sunny sky."
Computer model: "A man flying through the air while riding a snowboard."

How ?

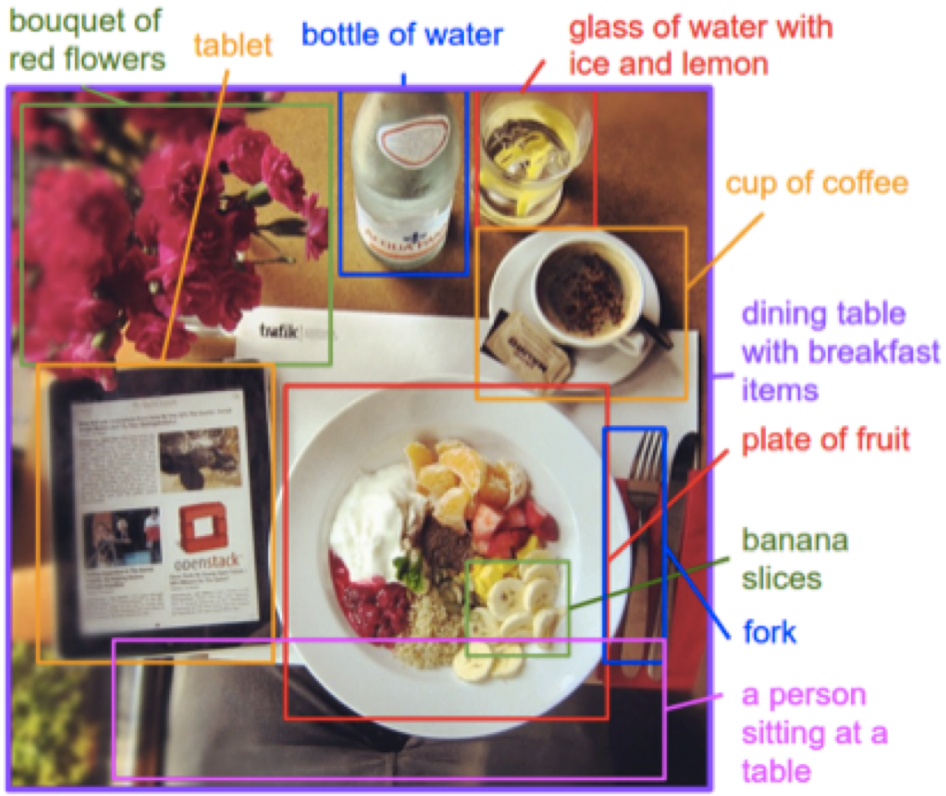


Figure 1. Our model generates free-form natural language descriptions of image regions.

Deep Learning used for Captioning

- Some of our automatic captions …

#990
a baseball player holding a bat on a field

#1599
a white cat sitting on top of a table

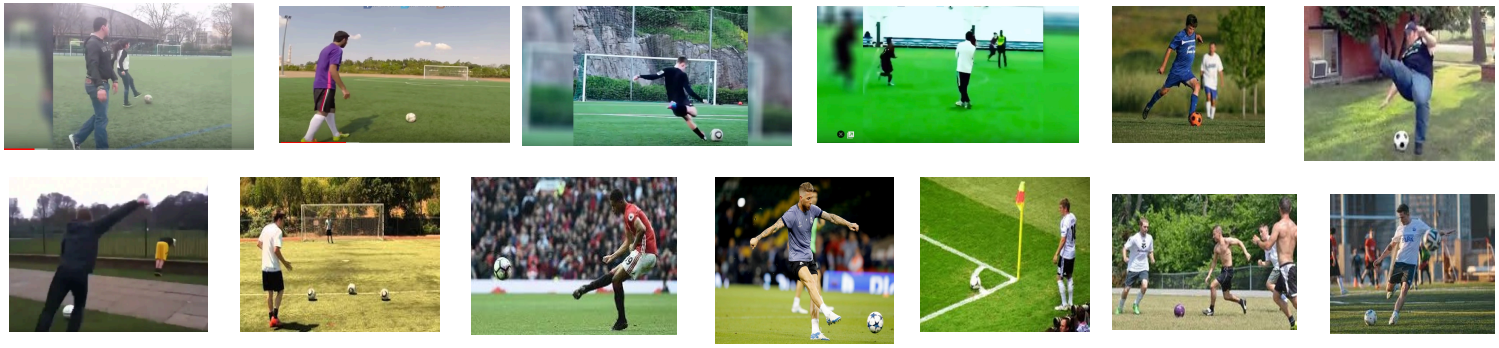
#603
a green truck is parked on a street

#1695
a person riding a bike down a street

DCU

Revealed Bias in Training Data

- We have many videos of men playing soccer



- … all manually captioned accordingly, used as training data, but ...

Google introduces machine learning analysis tool to combat AI bias

Sep 13, 2018 | Chris Burt

CATEGORIES Biometric R&D | Biometrics News

Google has unveiled a bias-detection feature for its TensorFlow machine learning web application, dubbed the What-If Tool, in a blog post.

BBC Sign in News Sport Weather Shop Earth Travel M

NEWS

Home Video World UK Business Tech Science Stories Entertainment & Arts

Technology

IBM launches tool aimed at detecting AI bias

By Zoe Kleinman
Technology reporter, BBC News

19 September 2018 Share

Algorithms and machine-learning could increasingly be used to calculate things like car insurance

IBM launches cloud tool to detect AI bias and explain automated decisions

Natasha Lomas @riptari / 1 day ago

IBM has launched a software service that scans AI systems as they work in order to detect bias and provide explanations for the automated decisions being made — a degree of transparency that may be necessary for compliance purposes not just a company's own due diligence.

The new trust and transparency system runs on the IBM cloud and works with models built from what IBM bills as a wide variety of popular machine learning frameworks and AI-build environments — including its own

Bloomberg the Company & Its Products | Bloomberg Anywhere Remote Login | Bloomberg Terminal Demo Request

Bloomberg

Technology

Accenture Unveils Tool to Help Companies Insure Their AI Is Fair

By Jeremy Kahn
June 13, 2018, 6:00 AM GMT+1

ACCENTURE AI TESTING SERVICES

Powered by our Teach and Test Methodology

accenturetechnology

ARTIFICIAL INTELLIGENCE

Microsoft Announces Tool To Catch Biased AI Because We Keep Making Biased AI

Flagging prejudiced algorithms won't keep them from being made in the first place.

Dan Robitzski | May 25th 2018

DCU

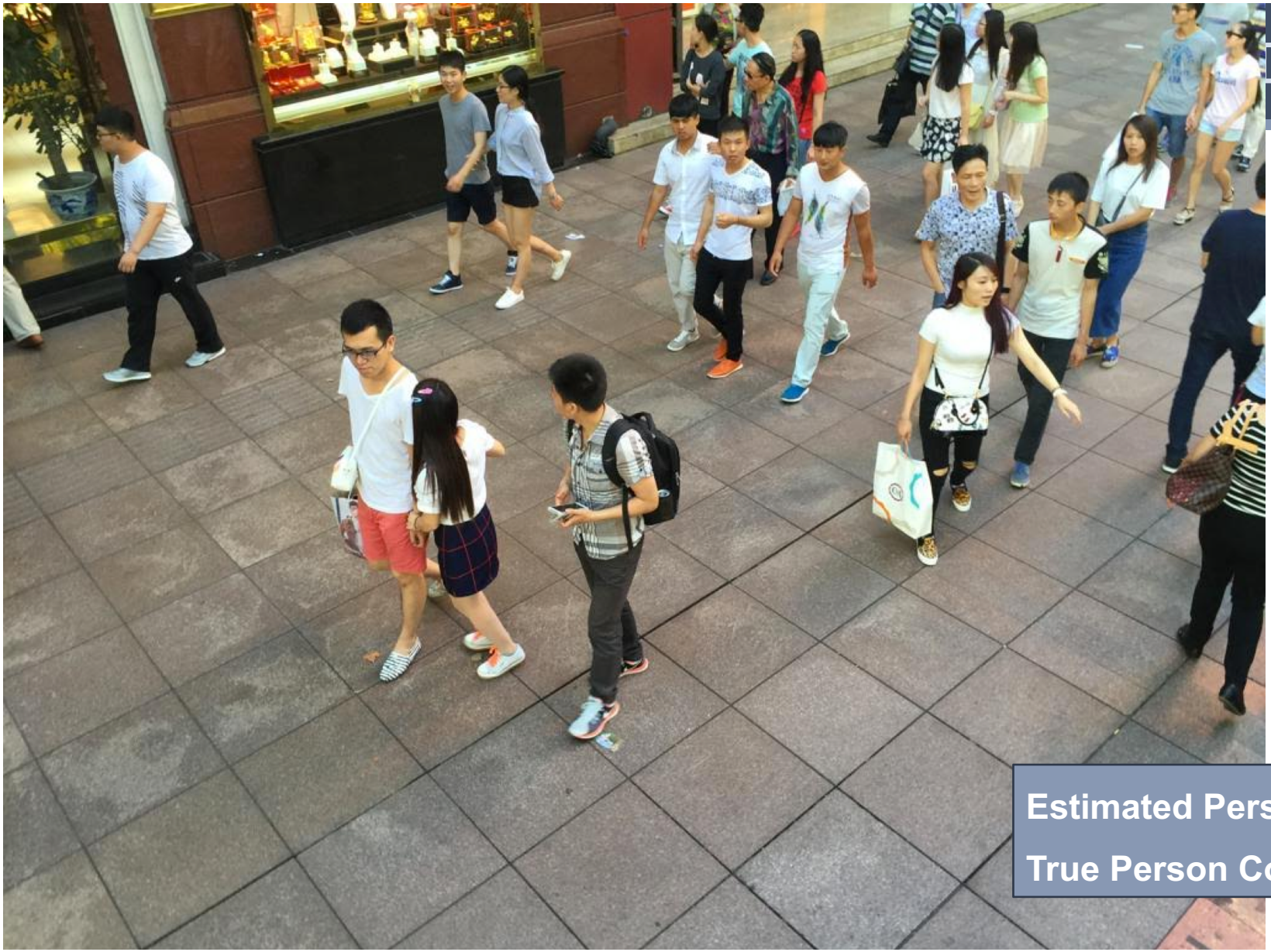
The World Loves Deep Learning

- With everybody “doing AI”, everything is “AI-based”, everybody loves deep learning … back to that in Part 2

- Meanwhile in MM research we use DL for other MM tasks besides tagging and captioning …

Datasets

Datasets



DCU

atasets

atasets

asets

asets

Estimated Person Count : 1544

True Person Count: 1566

Long-term video memorability scores

0.727 0.273

MediaEval Benchmark

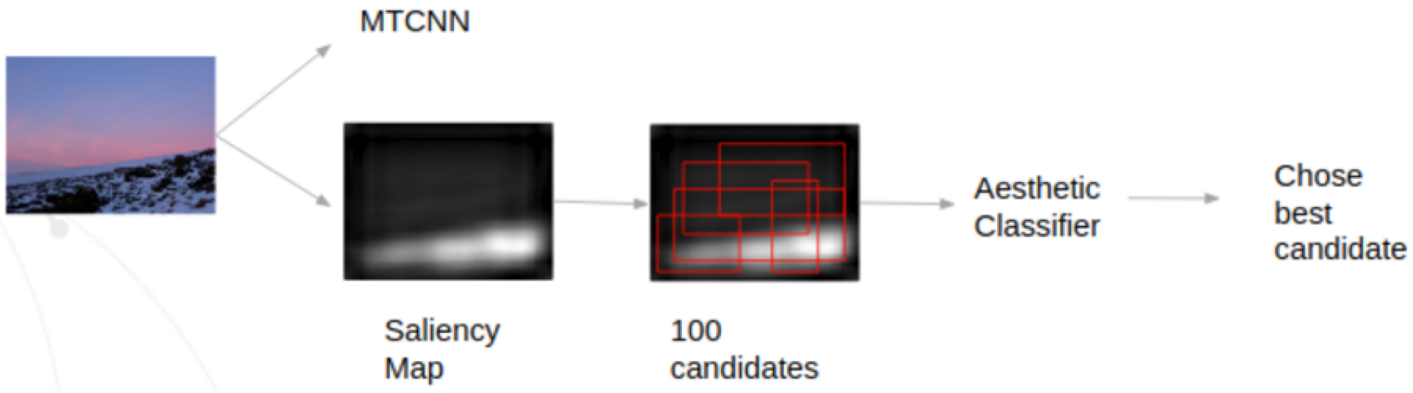
MediaEval Benchmarking Initiative for Multimedia Evaluation

Image Aesthetics

Top positive results

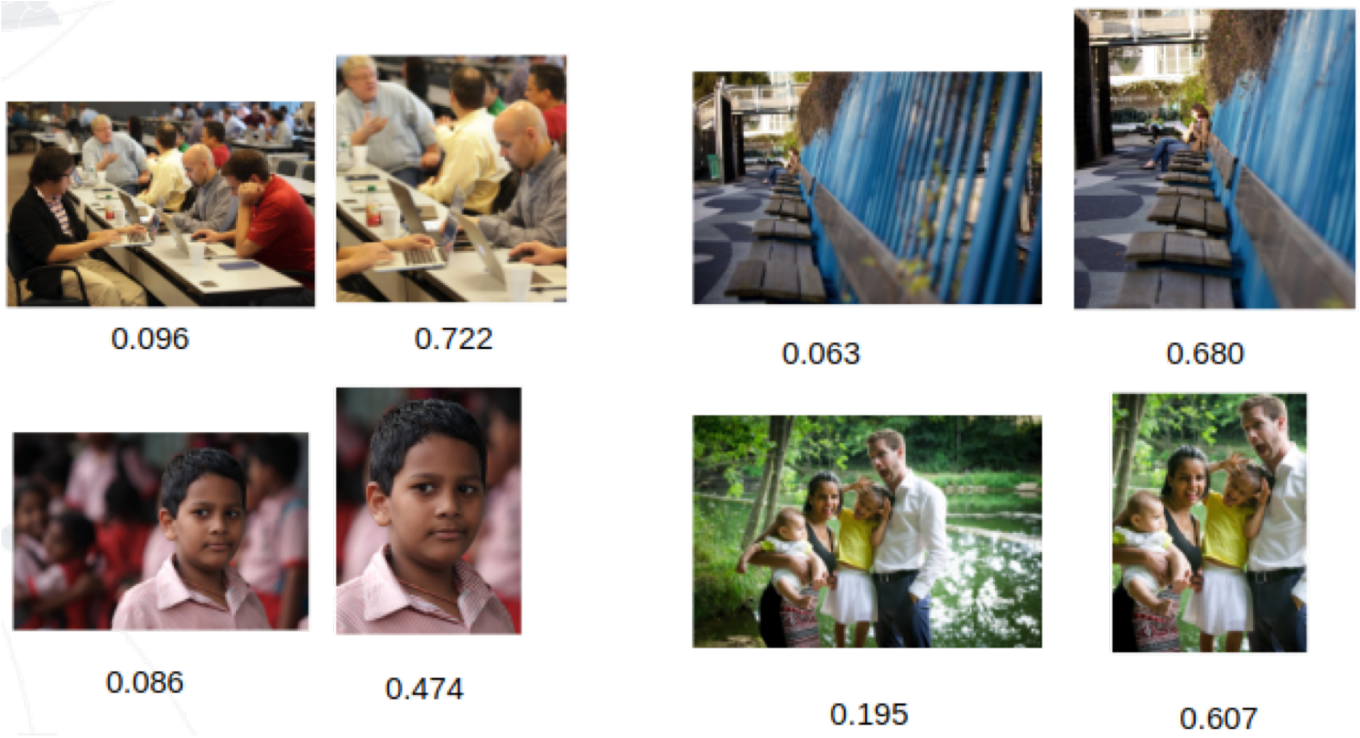
Top negative results

We push this more .. good images from mediocre ones ?



Uses our best-in-class salience detection

Our salience-based aesthetic auto-cropping



DCU

Never enough data …

- All of these content-based MM apps …
  - captions, person-counting, memorability, aesthetics, salience, auto-cropping

  based on deep learning
- All have challenges of model architecture, training time .. and **training data** … both **volume** and **bias**
- We use "tricks" to do data augmentation
- Our greatest trick is GANs … in part 2

End of part 1

Generative Forms of Multimedia Content

Part 2

Alan F. Smeaton

Dublin City University

Invited Keynote talk presented at IEEE Conference on Multimedia and Expo (ICME)

7th July 2020

A World Leading SFI Research Centre

Science Foundation Ireland SFI For what's next

Insight

SFI RESEARCH CENTRE FOR DATA ANALYTICS

HOST INSTITUTIONS

DCU

NUI Galway OÉ Gaillimh

UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh

UCD Dublin

PARTNER INSTITUTIONS

Maynooth University National University of Ireland Maynooth

Tyndall National Institute

Trinity College Dublin The University of Dublin

University of Limerick Ollscoil Luimnigh

FUNDED BY:

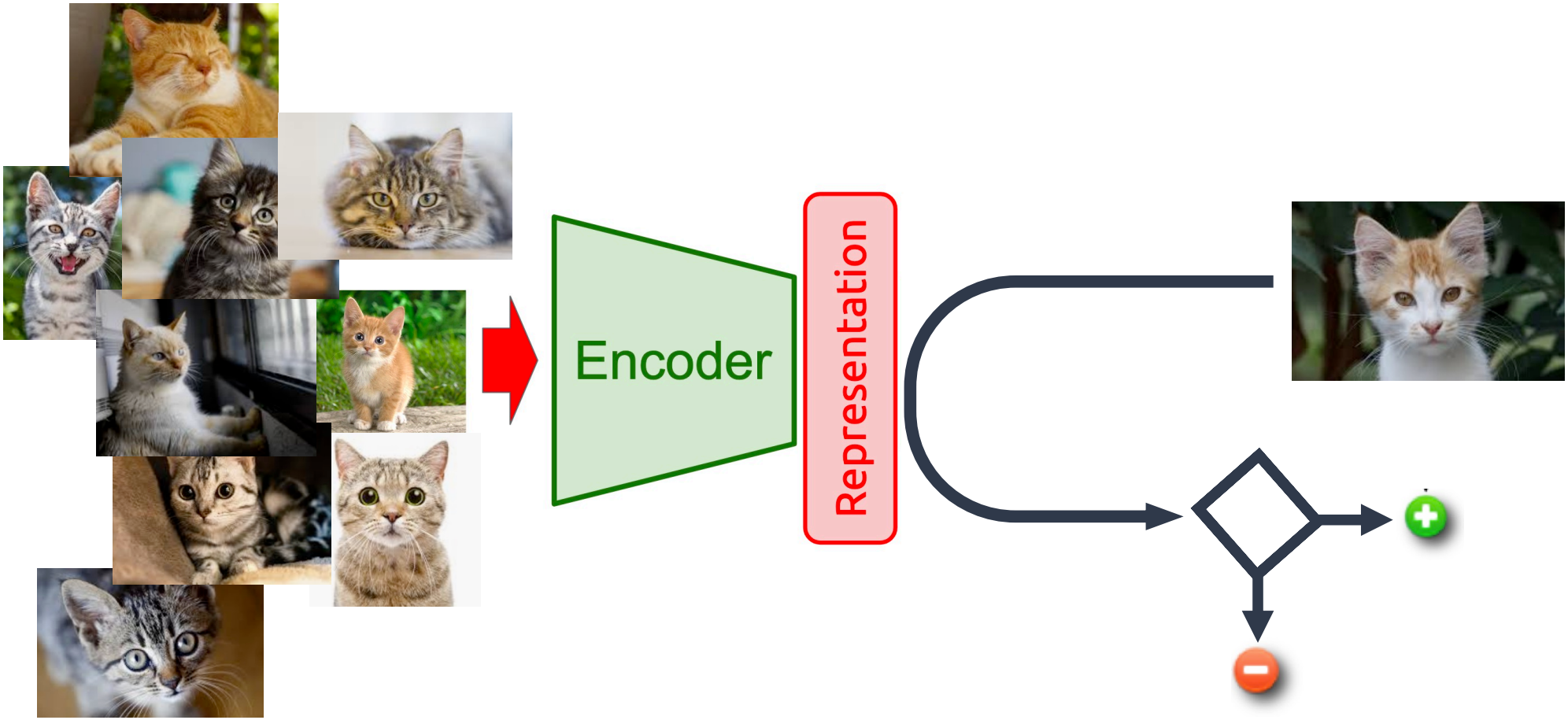
European Union European Regional Development Fund

Science Foundation Ireland SFI For what's next

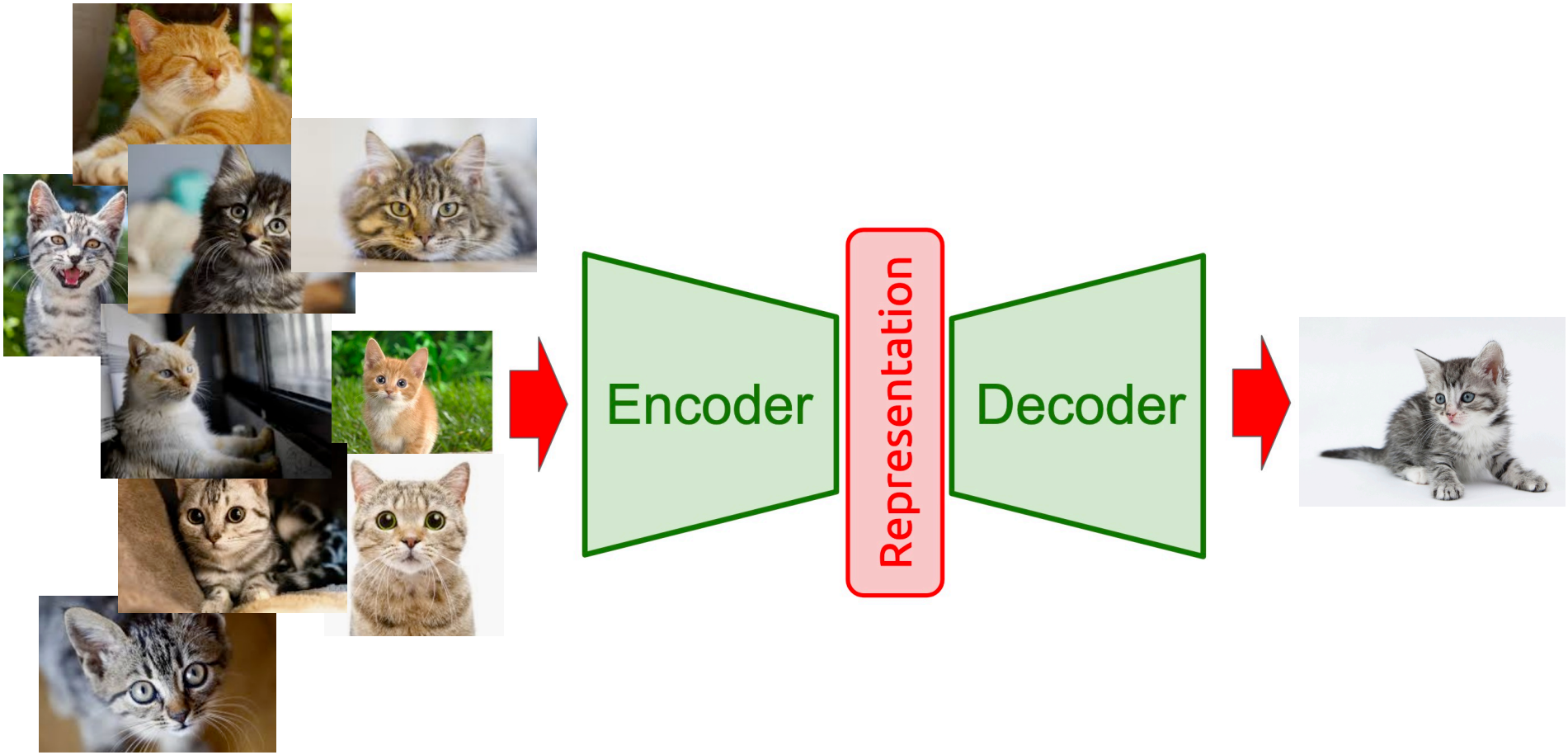
ML has issues …

- Lots of attention on deep learning, with issues like
  - model building, adjusting hyper-parameters
  - new architectures like capsule networks
  - replicating the (human) brain's neural structures beyond connections, like how neurotransmitters work
  - computation costs, from GPUs to custom
  - explainable AI / ML, justifying outputs, data bias
- Successful computer vision/MM applications need training data -> data augmentation

Conventional Modeling

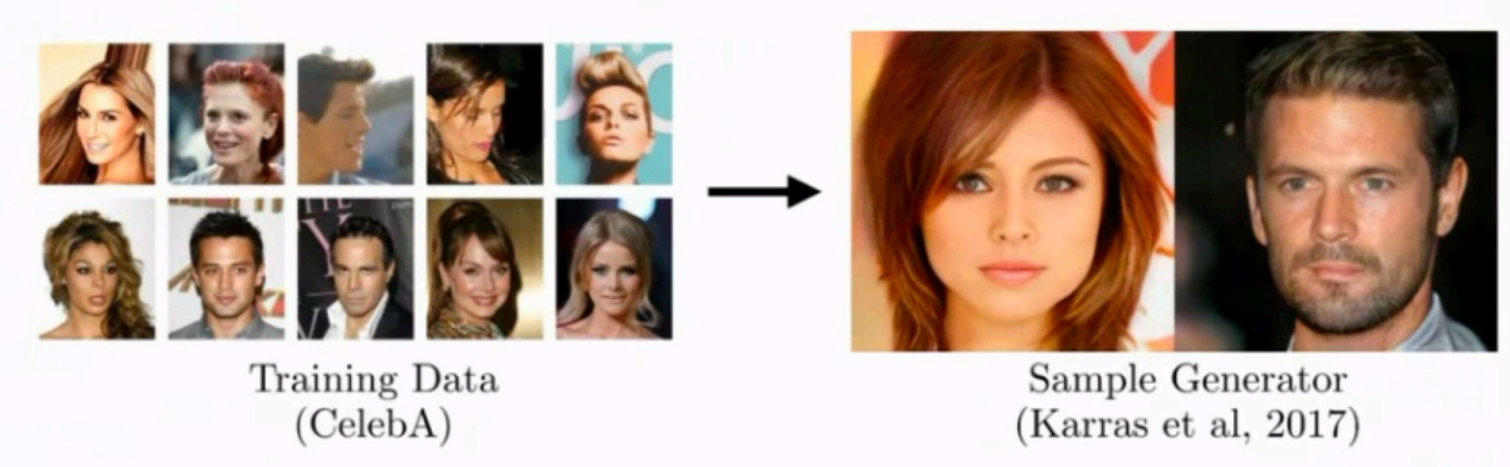


Generative Modeling



Generative Modeling

- **Generates** rather than **classifies** or **predicts** (previous applications of ML) .. learns probability distributions from some training data and then generates or synthesises new (images) from those distributions



What are GANs

- In 2014 GANs were introduced:
  - Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014.
- GANs quality improved … rapidly … on faces, which are easy

Image credit Ian Goodfellow

Beyond Faces?

- More challenging, but other objects now achievable

Image credit Ian Goodfellow

Reviews & categorises 322 papers

Generative Adversarial Networks: A Survey and Taxonomy

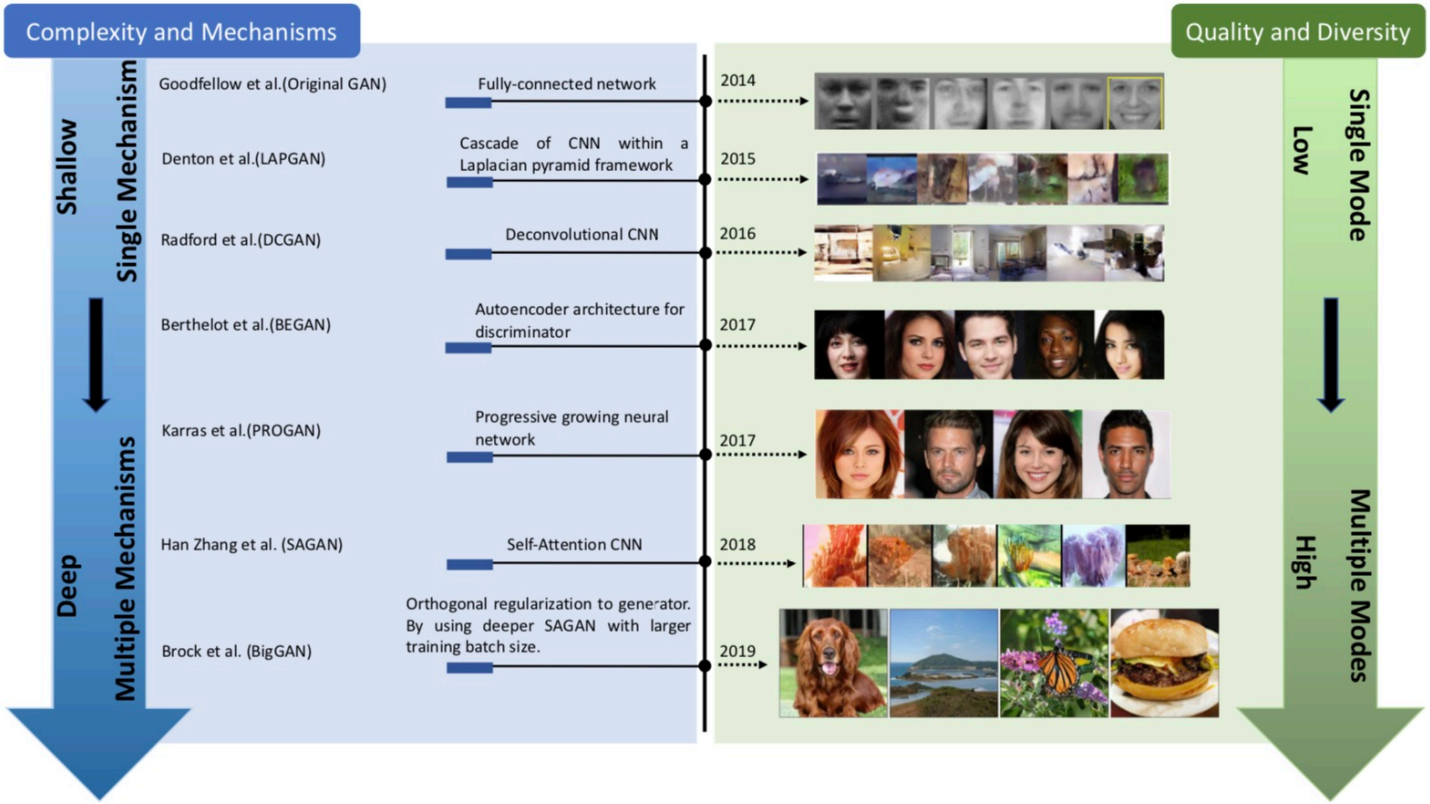
Zhengwei Wang, Qi She, Tomás E. Ward

Abstract—

Generative adversarial networks (GANs) have been extensively studied in the past few years. Arguably the revolutionary techniques are in the area of computer vision such as plausible image generation, image to image translation, facial attribute manipulation and similar domains. Despite the significant success achieved in the computer vision field, applying GANs to real-world problems still poses significant challenges, three of which we focus on here: (1) High quality image generation; (2) Diverse image generation; and (3) Stable training. Through an in-depth review of GAN-related research in the literature, we provide an account of the architecture-variants and loss-variants, which have been proposed to handle these three challenges from two perspectives. We propose loss-variants and architecture-variants for classifying the most popular GANs, and discuss the potential improvements with focusing on these two aspects. While several reviews for GANs have been presented to date, none have focused on the review of GAN-variants based on their handling the challenges mentioned above. In this paper, we review and critically discuss 7 architecture-variant GANs and 9 loss-variant GANs for remedying those three challenges. The objective of this review is to provide an insight on the footprint that current GANs research focuses on the performance improvement. Code related to GAN-variants studied in this work is summarized on *https://github.com/sheqi/GAN_Review*.

Index Terms—Generative Adversarial Networks, Computer Vision, Architecture-variants, Loss-variants, Stable Training.

Reviews & categorises 322 papers



GAN applications

- GANs used for generating realistic examples of paintings, videos, images, text, 3D models (for replacement teeth !), DNA sequences …

- *"… plausible image generation, image to image translation, facial attribute manipulation and similar domains"*

Deoldify B&W Images

- Training data is colour images of similar scenes
- Multiple level representations allow realistic shadows and textures

Jason Antic, with thanks to John Breslin at Insight @ NUI Galway

Deoldify B&W Images

Jason Antic, with thanks to John Breslin at Insight @ NUI Galway

Deoldify B&W Images

Jason Antic, with thanks to John Breslin at Insight @ NUI Galway

Paintings …

GAN-generated "painting" called "Edmond de Belamy"

Paintings …

GAN-generated "painting" called "Edmond de Belamy"

Sold for $435,000 at auction in Christies in 2018

Video Style Transfer

Thanks to Enric Moreau

Bill Hader / Al Pacino (Scarface version)

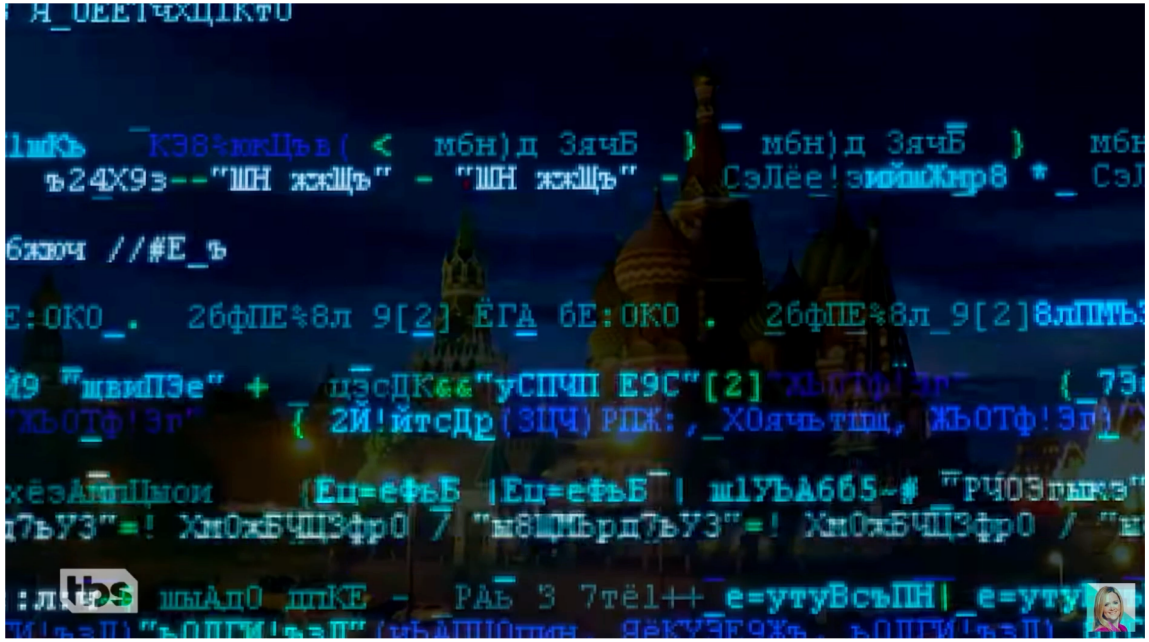
Dali is Back !

Dali museum, St Petersburg, Florida … 125 such videos, 45 minutes total, 1,000 hours of training time from 6,000 images https://www.youtube.com/watch?v=MZ2X-fSIPSU

Beckham and Malaria

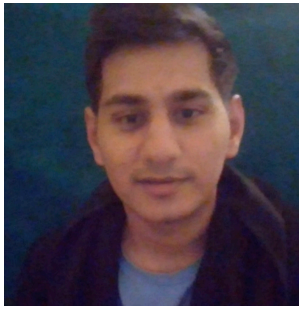
https://www.youtube.com/watch?v=QiiSAvKJIHo

Political Deepfakes ?



Video from Samantha Bee at TBC

Deepfakes are easy

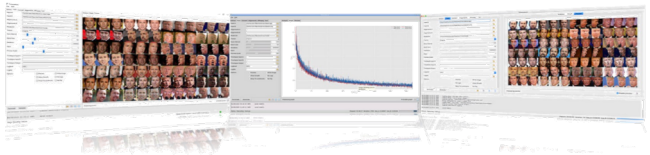


faceswap

WELCOME

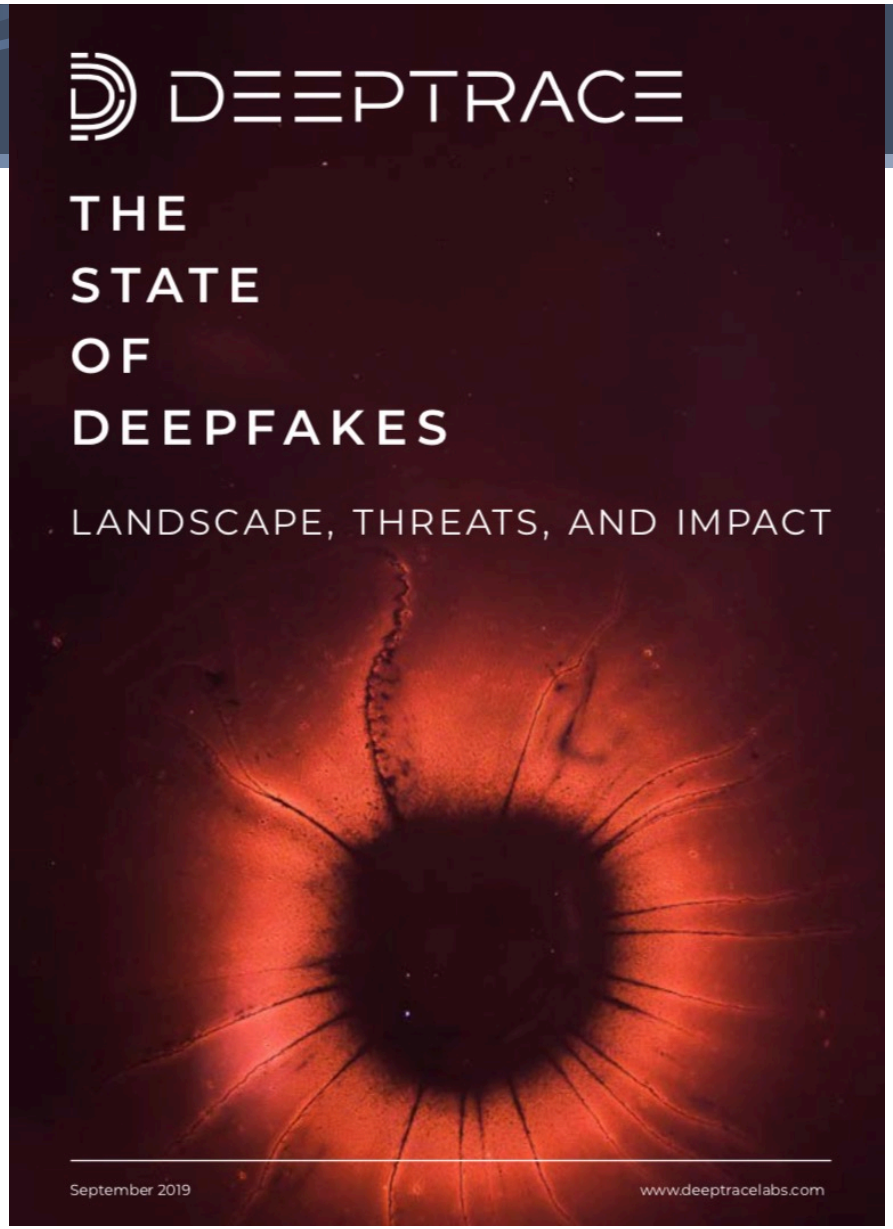
Jun, 30 2019 (Last Update: Thu, 15 Aug 2019)

Faceswap is the leading free and Open Source multi-platform Deepakes software.



Powered by Tensorflow, Keras and Python; Faceswap will run on Windows, macOS and Linux.

DCU



Total number of deepfake videos online

14,678

percentage of deepfake videos online by pornographic and non-pornographic content

96%

4%

Total number of video views across top four dedicated deepfake pornography websites

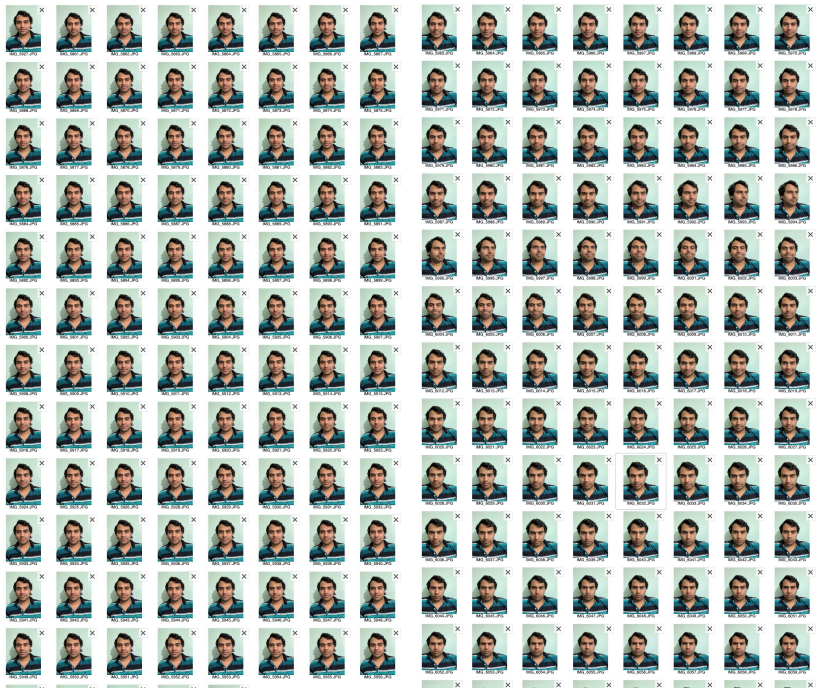
134,364,438

Deepfake Detection

- Eulerian Video Magnification from MIT, @SIGGraph 2012 and ACM ToG
- Magnifies imperceptual motion and movement in original videos

Minimal Deepfakes

- Use case ?
- How much facial variation is needed for realistic deepfakes ?
- Measure quality of GAN output by Inception Score or Fréchet Inception Distance or Synthetic Neuroscore
- Use collections of training images



Is this AI?

- What we call AI, almost entirely ML, works when ..
  - **mimics** human decision-making that doesn't change over time, no evolution
  - well-defined **inputs**, well-defined **outputs**
  - **large** digital data sets as a basis for training
  - no long chains of complicated **logic** or inference or reasoning
  - not much **background knowledge**,
  - don't need to **explain** the output, can tolerate some error
- We started with definition(s) of AI

AI Definitions

- AI definition is old, changes as tech developed
- Several dictionaries define AI as …
  (taken from Forbes Magazine)
  - "The theory and development of computer systems able to perform tasks **normally requiring human intelligence**, such as visual perception, speech recognition, decision-making, and translation between languages."
  - "A branch of computer science dealing with the simulation of **intelligent behaviour** in computers."
  - "The capability of a machine to **imitate intelligent human** behaviour."
  - "the ability of a digital computer or computer-controlled robot to perform **tasks** commonly **associated with intelligent** beings."
  - "The field of computer science dedicated to solving cognitive problems commonly **associated with human intelligence**, such as learning, problem solving, and pattern recognition."

Conclusions

- That was a roller-coaster journey through AI, ML, deep learning, computer vision, tagging, captions, data insufficiency, data bias, data augmentation, generative ML, stunning examples and deepfakes
- ML, which underpins almost all of current AI, has its origins in CV but everything here on ML has mapping into other problem domains .. Fintech, medicine, NLP, education, entertainment, anywhere you see “AI” used
- But almost every example .. is *just* ML .. and ML is *just* replication .. and replication is only part of the definition of AI
- So it is Artificial Intelligence, but not as we know it (Jim!)

Finally, Thanks to …

- … to PhD students and graduates, to Postdoctoral Researchers, Research Assistants, Masters students, colleagues and friends
- … to funding agencies, current ones are