



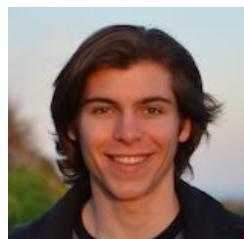
EPIC

Third International Workshop on Egocentric Perception, Interaction and Computing

Munich, Germany
9th September, 2018

PathGAN:

Visual Scanpath Prediction with Generative Adversarial Networks



Marc
Assens



Kevin
McGuinness



Noel E.
O'Connor



Xavier
Giro-i-Nieto



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH



Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Let's play a game!



The importance of visual attention







The importance of visual attention





The importance of visual attention







The importance of visual attention





The importance of visual attention





Why don't we see the changes?

We don't really see the whole image

We only focus on small specific regions: the **salient** parts

Human beings reliably attend to the same regions of images
when shown

What we perceive



Where we look



What we actually see





Why predicting **WHERE** humans will look?



Johanna Närväinen, Janne Laine, "Looking through their eyes" VTT Impulse 2016.





Can we predict **WHERE** humans will look?

Yes! We “just” need to collect data...





Can we predict WHERE humans will look?

Yes! We “just” need to collect data...

...and train a deep neural network !

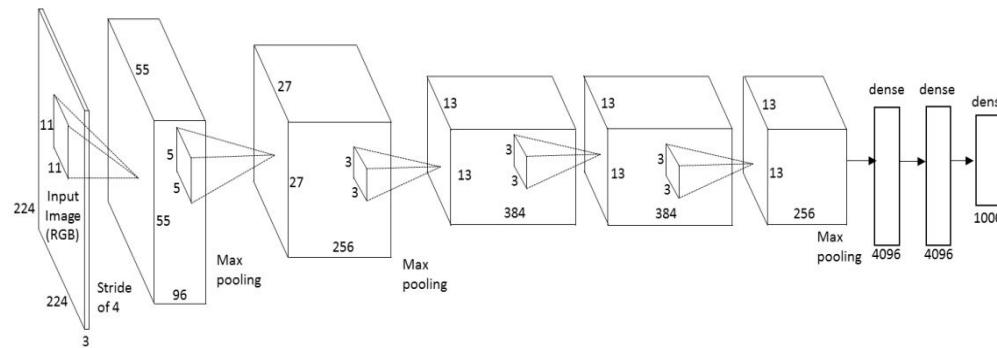
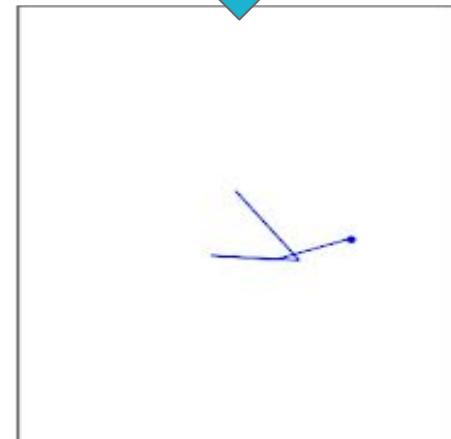
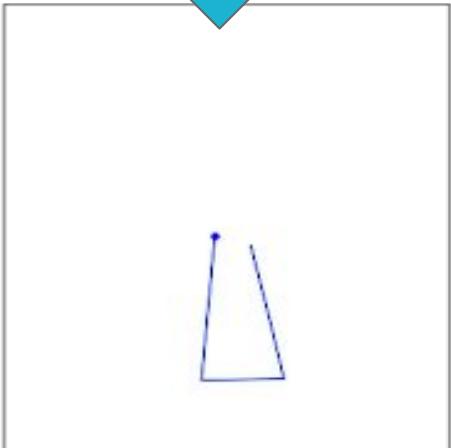




Image
(input)



Visual
scanpath
(output)





Our brief history of saliency models



Eva
Mohedano



Junting
Pan



Èric
Arazo



Marta
Coll



Panagiotis
Linardos



Cristian
Canton



Elisa
Sayrol



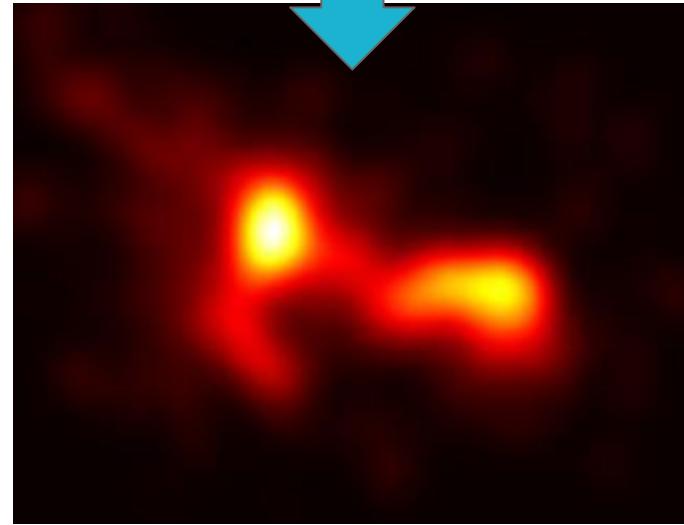
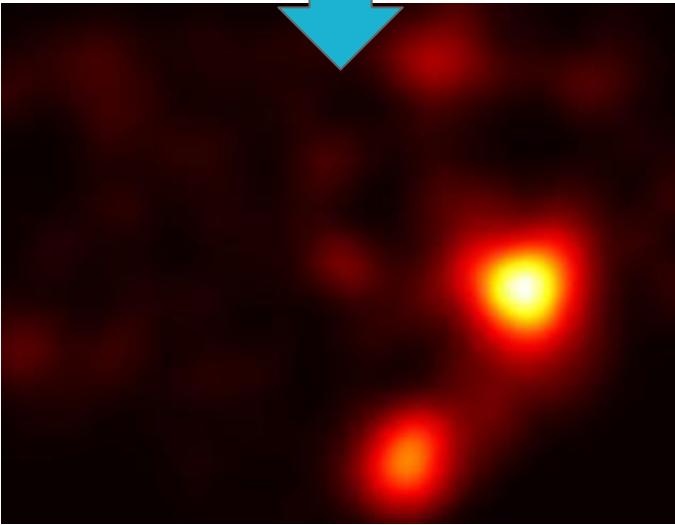
Jordi
Torres



Image
(input)



Saliency
map
(output)

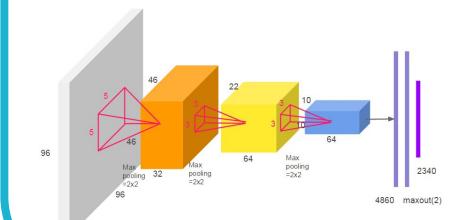


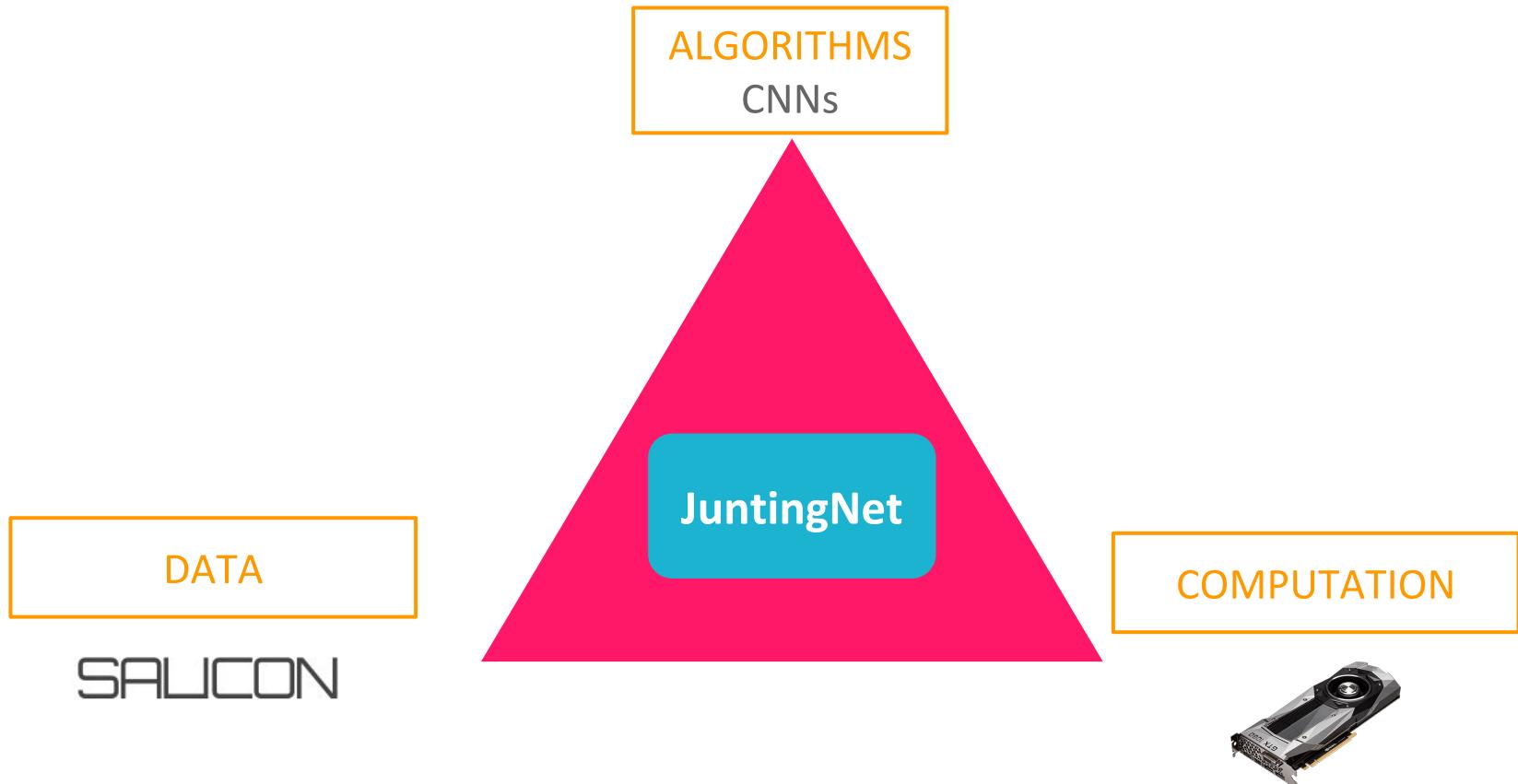


Saliency maps



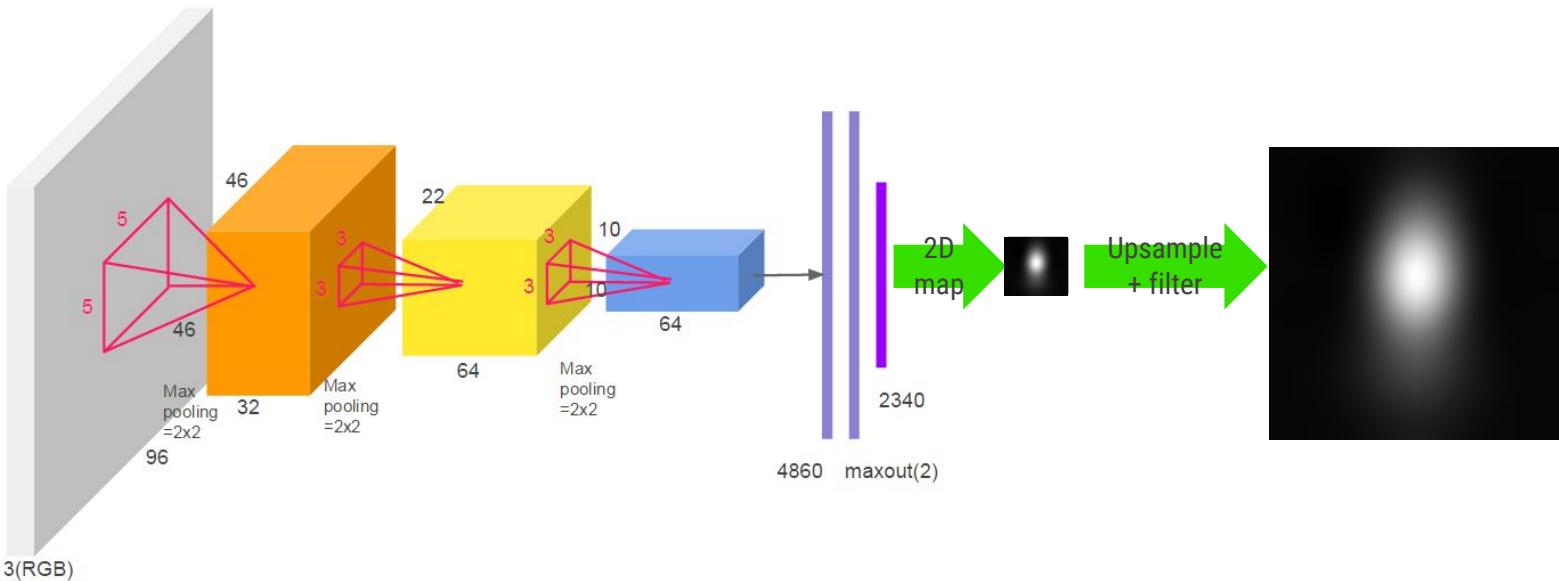
JuntingNet







JuntingNet

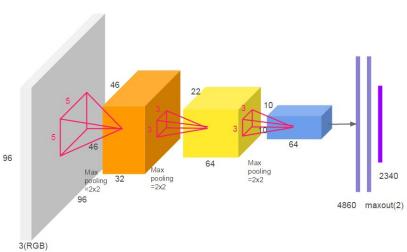




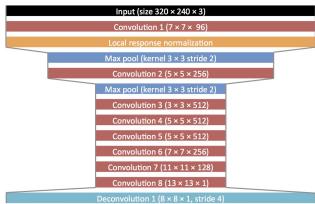
Saliency maps



JuntingNet



SalNet





ALGORITHMS

CNNs

SalNet

DATA

IMAGENET



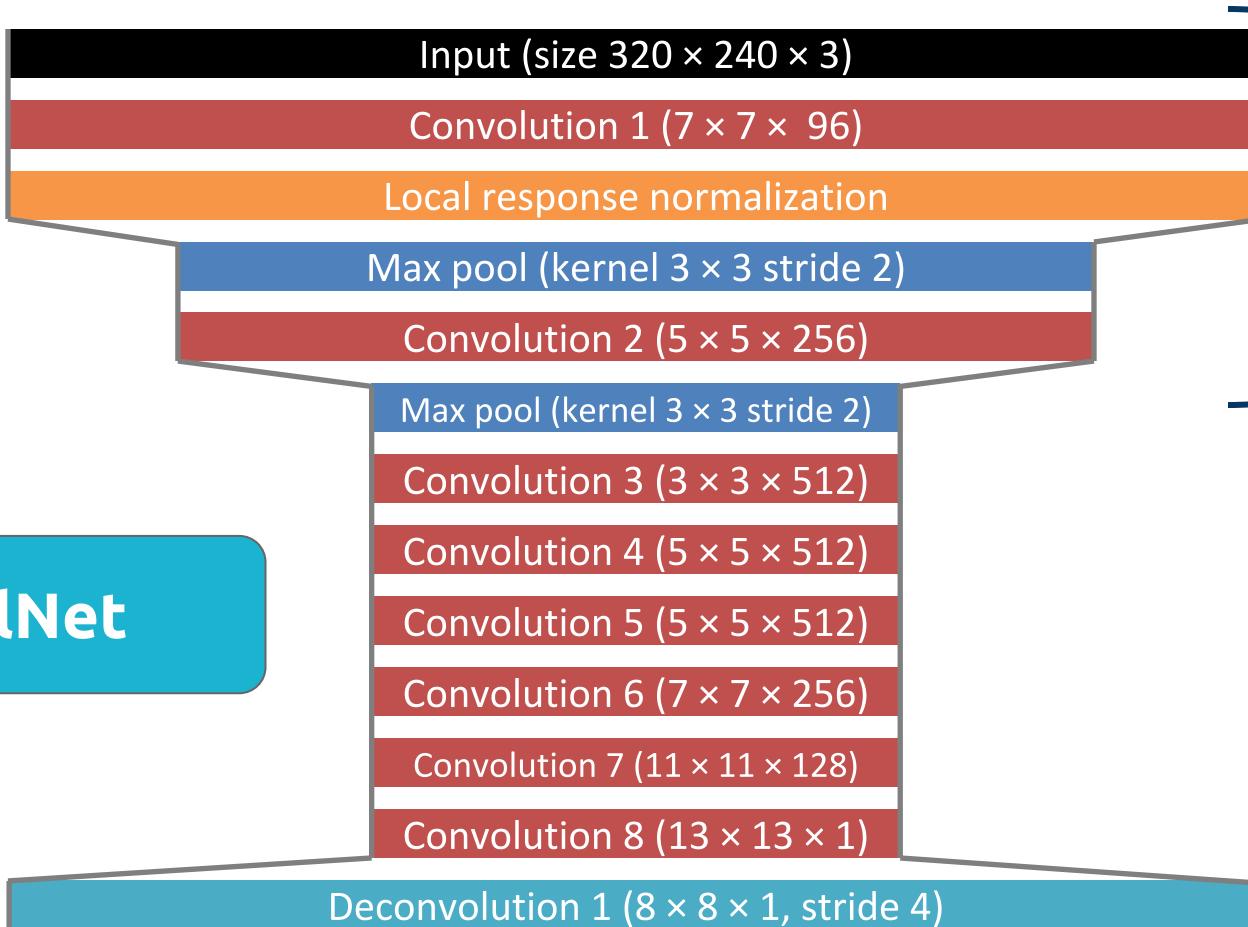
SALICON

COMPUTATION





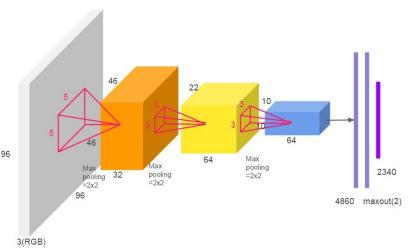
SalNet



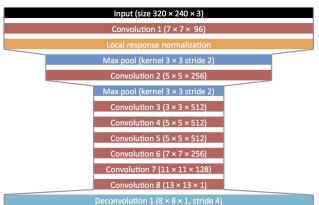
Saliency maps



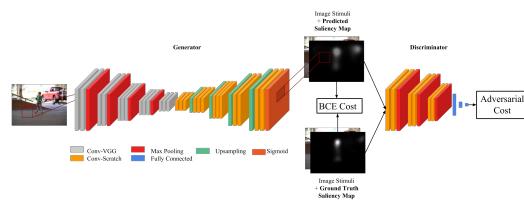
JuntingNet

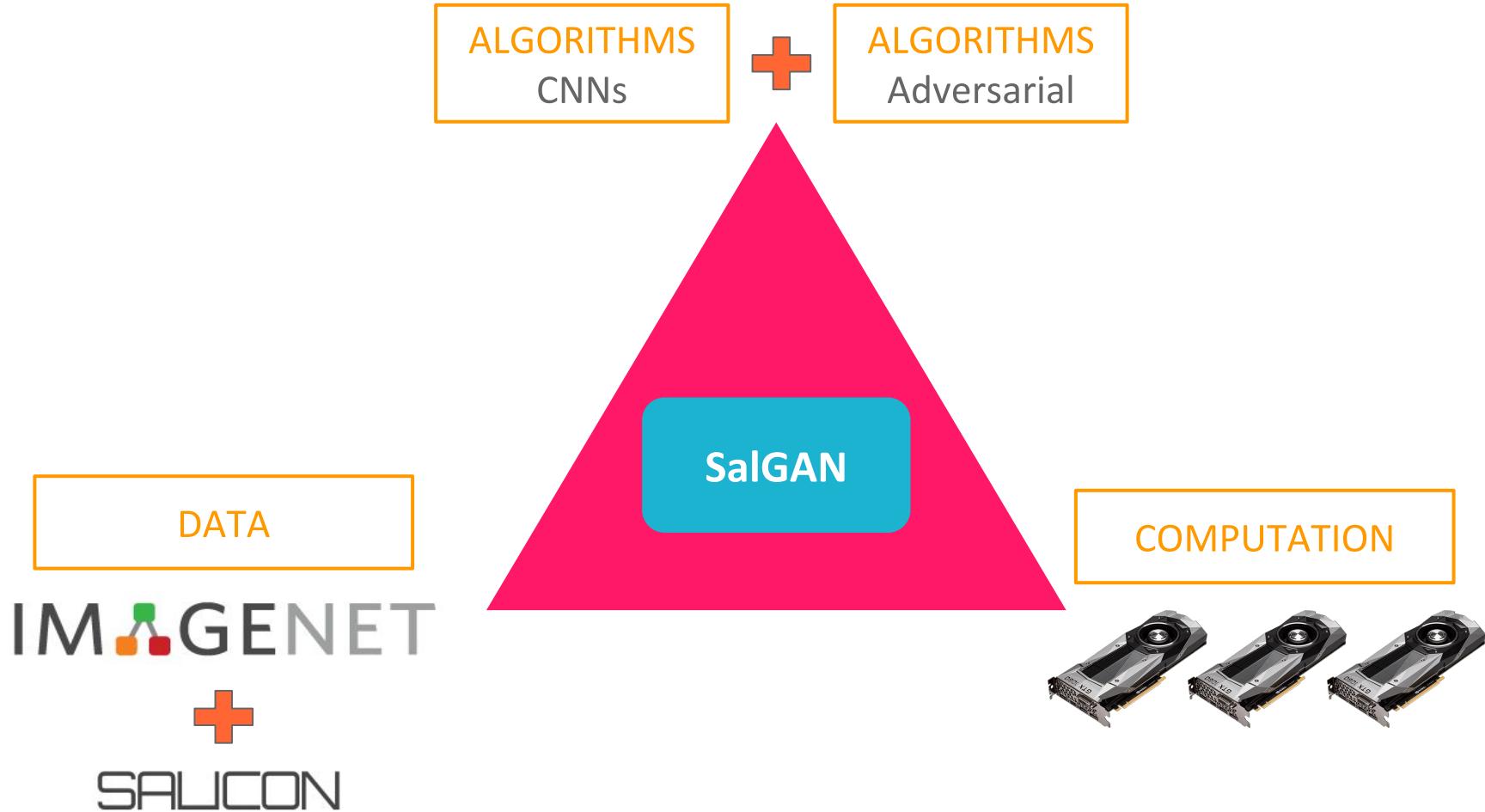


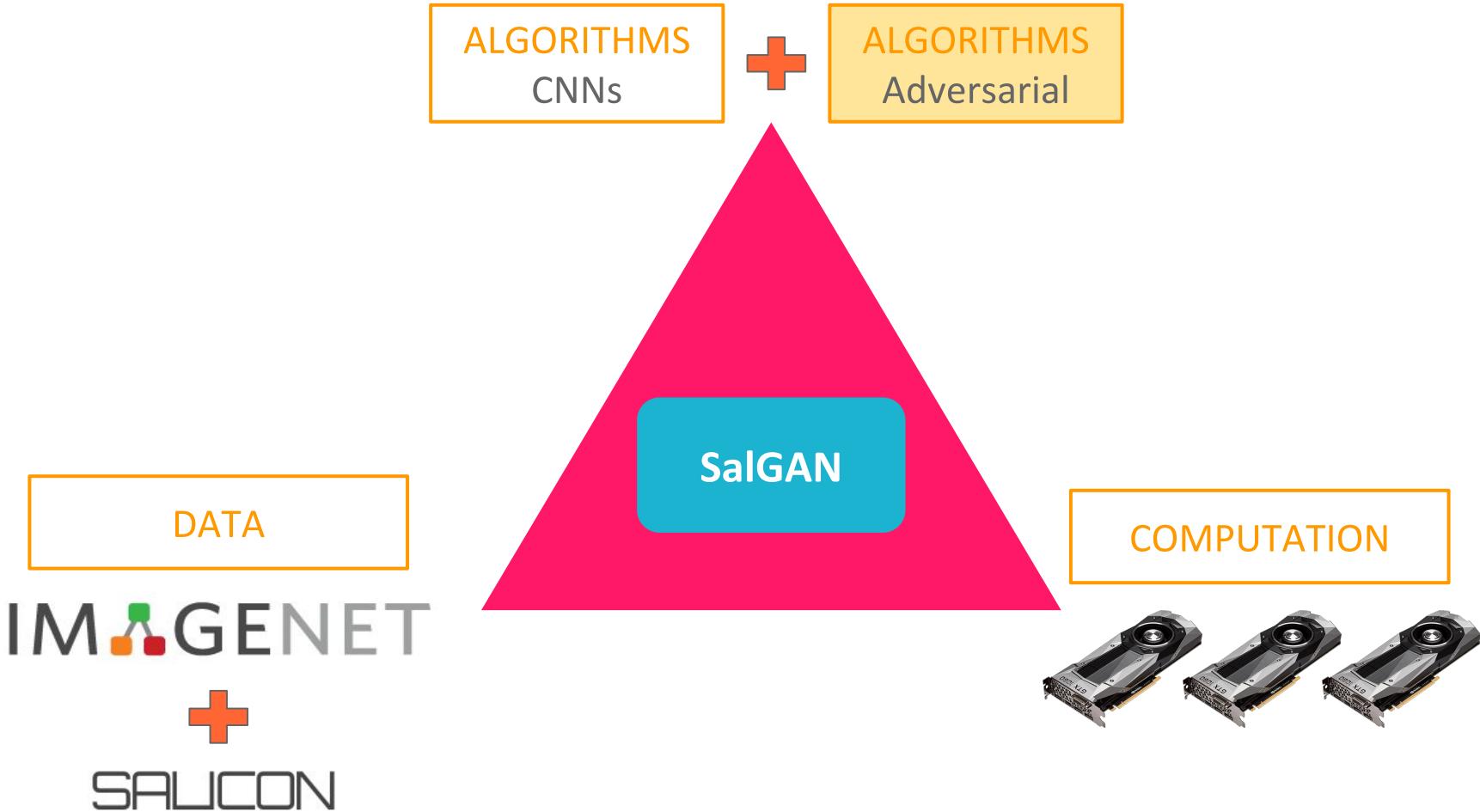
SalNet



SalGAN

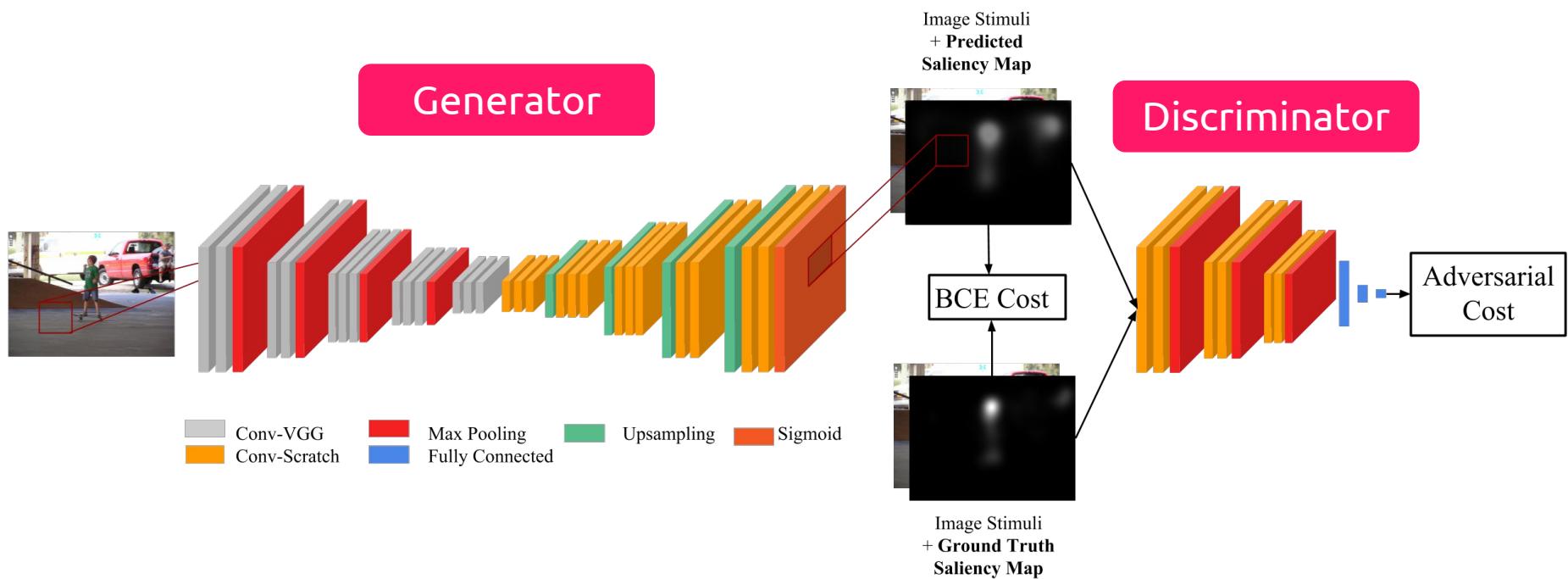








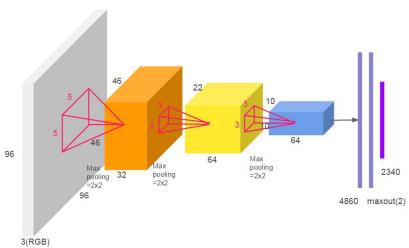
SalGAN



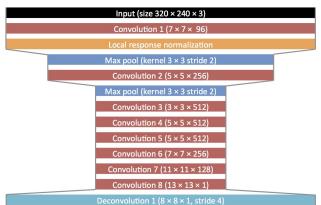
Saliency maps



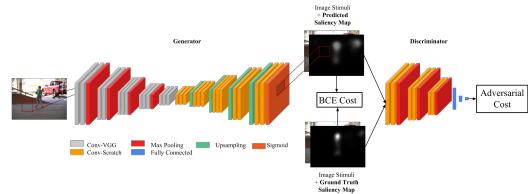
JuntingNet



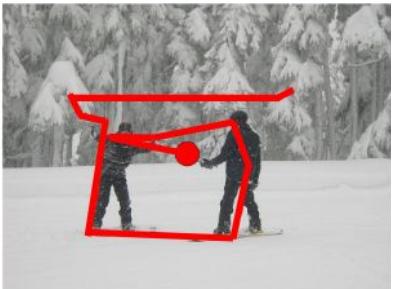
SalNet



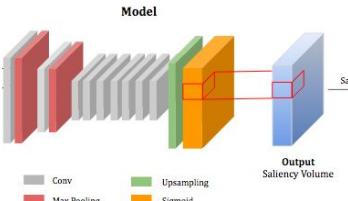
SalGAN

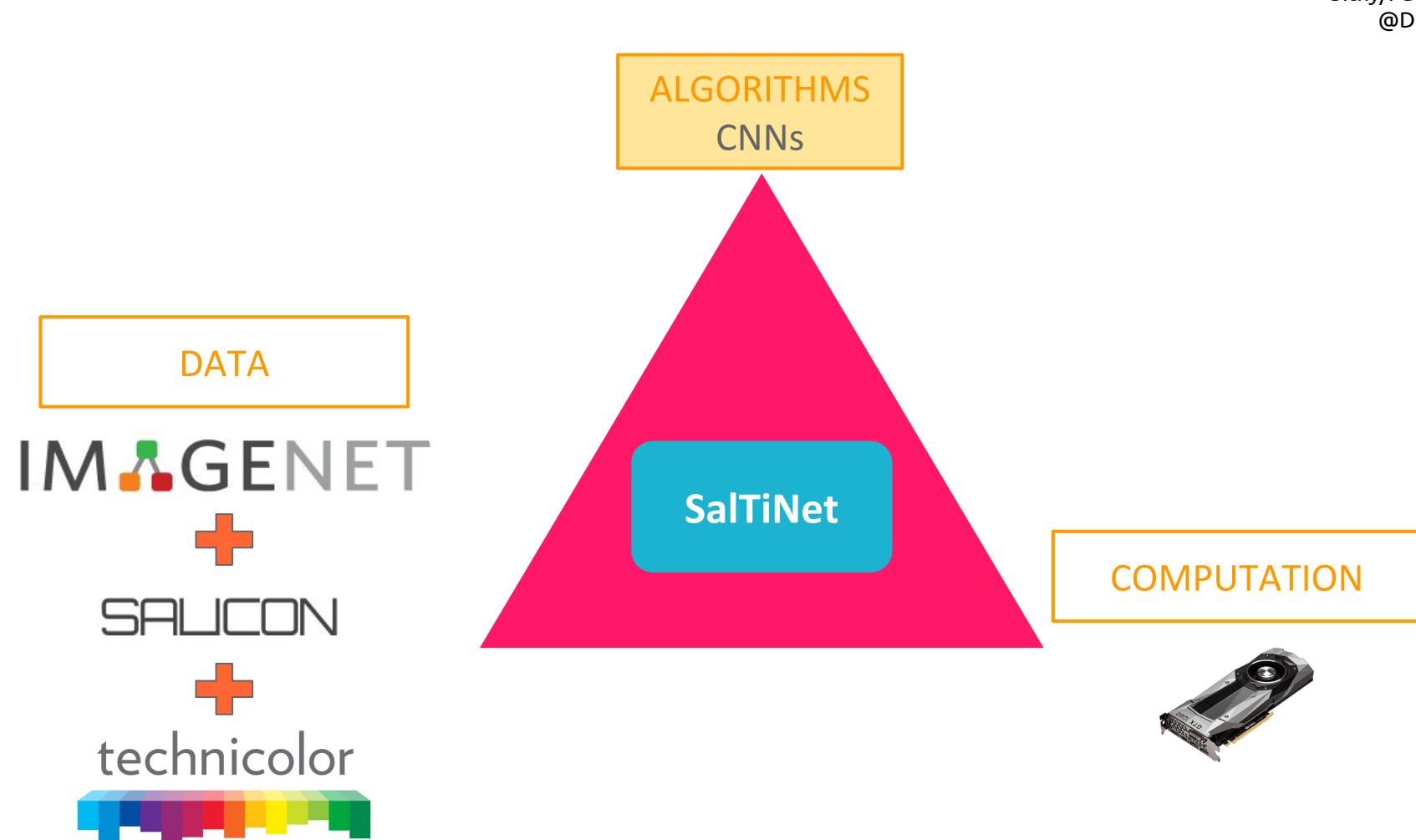


Scanpaths

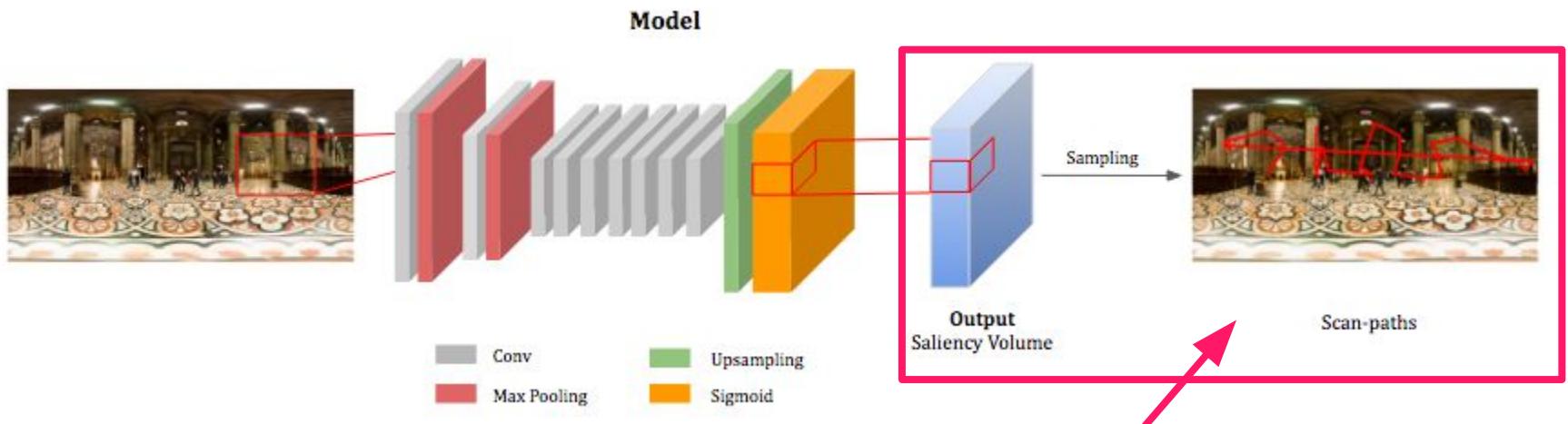


Saltnet





SalTiNet



Stochastic sampling

Winners of IEEE ICME 2017 Challenge



technicolor

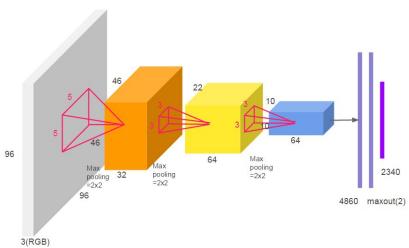



Marc Assens, Kevin McGuinness, Xavier Giro-i-Nieto and Noel E. O'Connor. "[SaltiNet: Scan-Path Prediction on 360 Degree Images Using Saliency Volumes.](#)" ICCVW 2017.

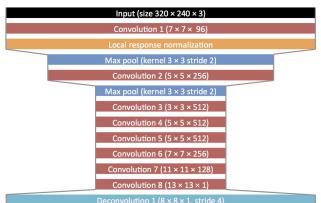
Saliency maps



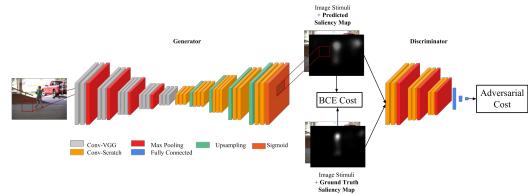
JuntingNet



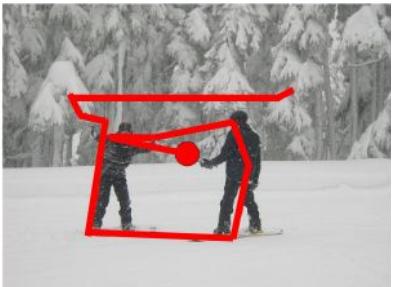
SalNet



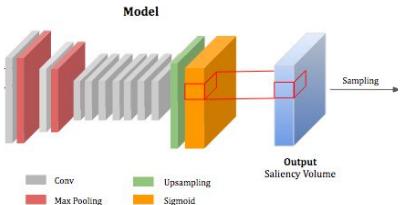
SalGAN



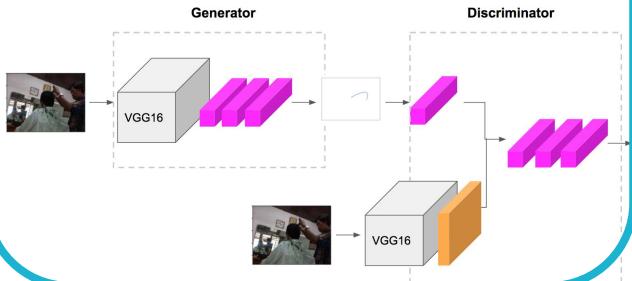
Scanpaths

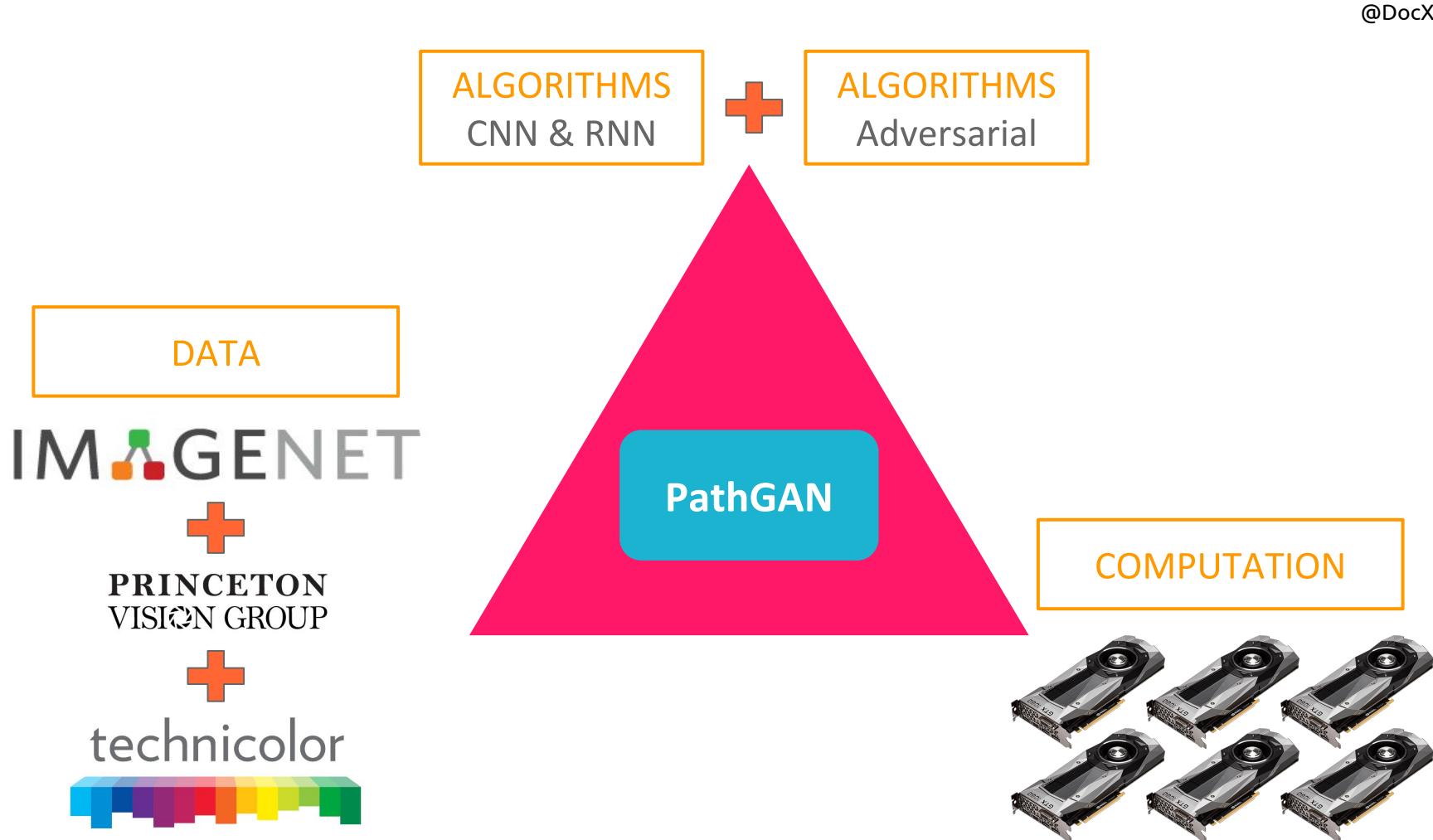


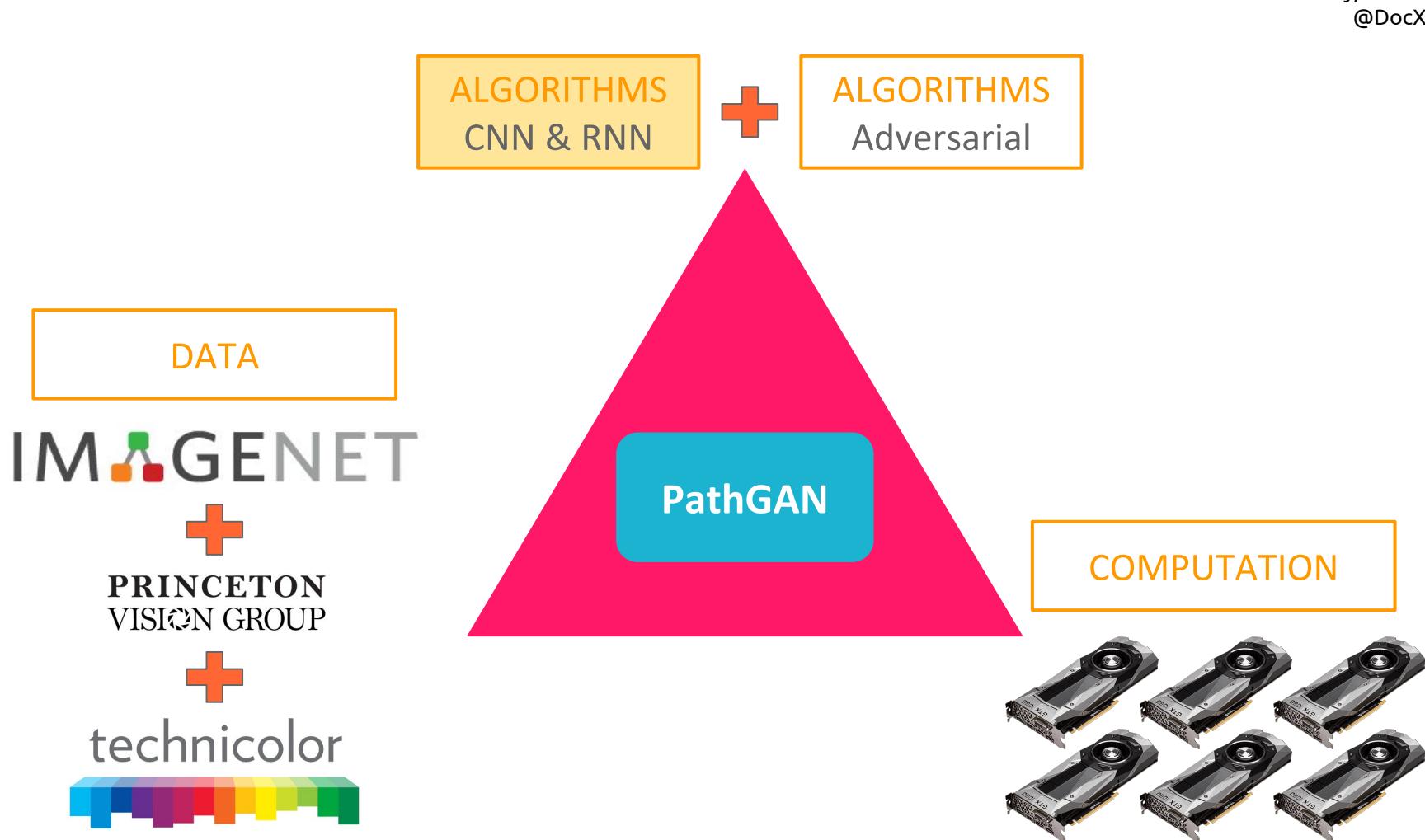
Saltnet



PathGAN









Challenge:

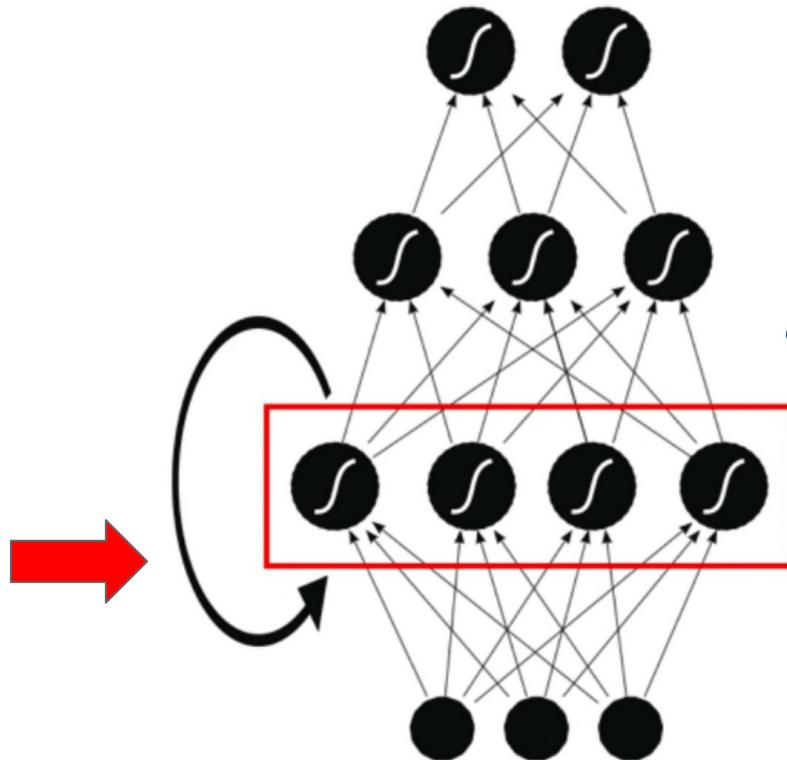
Create a model able to predict visual scanpaths

PathGAN solution:

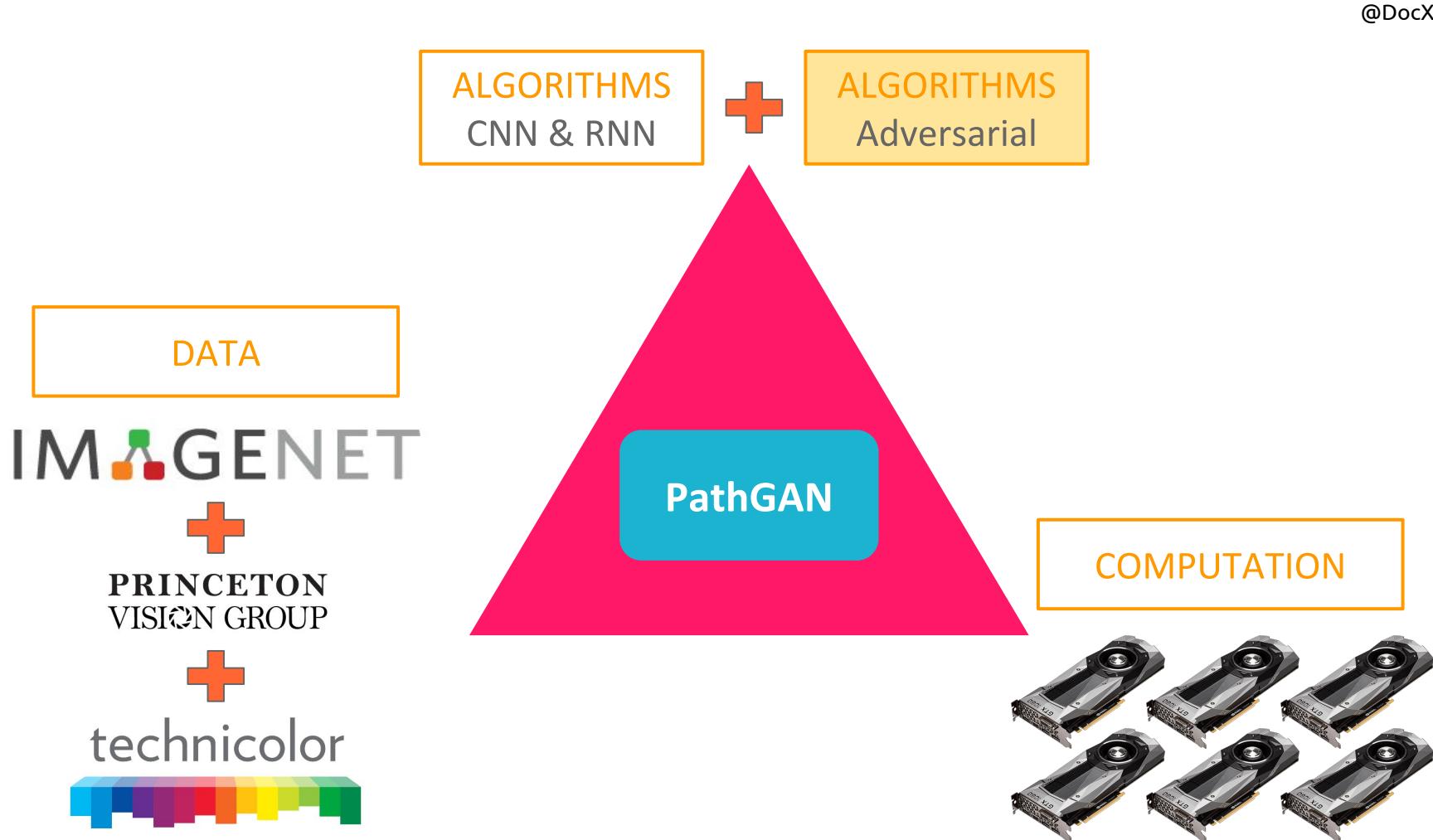
Generate a sequence of fixation points with

Recurrent Neural Networks (RNNs).

Recurrent Neural Network (RNN)



Hochreiter, Sepp, and Jürgen Schmidhuber. "[Long short-term memory.](#)" Neural computation 9, no. 8 (1997): 1735-1780.



TP 276100 HI 276100 FREE PLAY

KO

99

GENERATOR

DISCRIMINATOR

ROUND 1

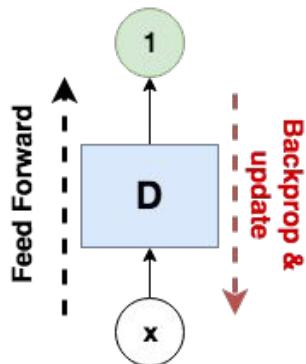
BATTLE 04





Adversarial training in a nutshell

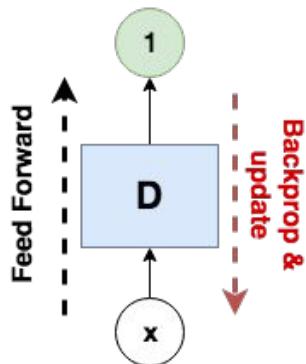
- Pick a real scanpath x from training set
- Show x to D and update weights to output 1 (real)





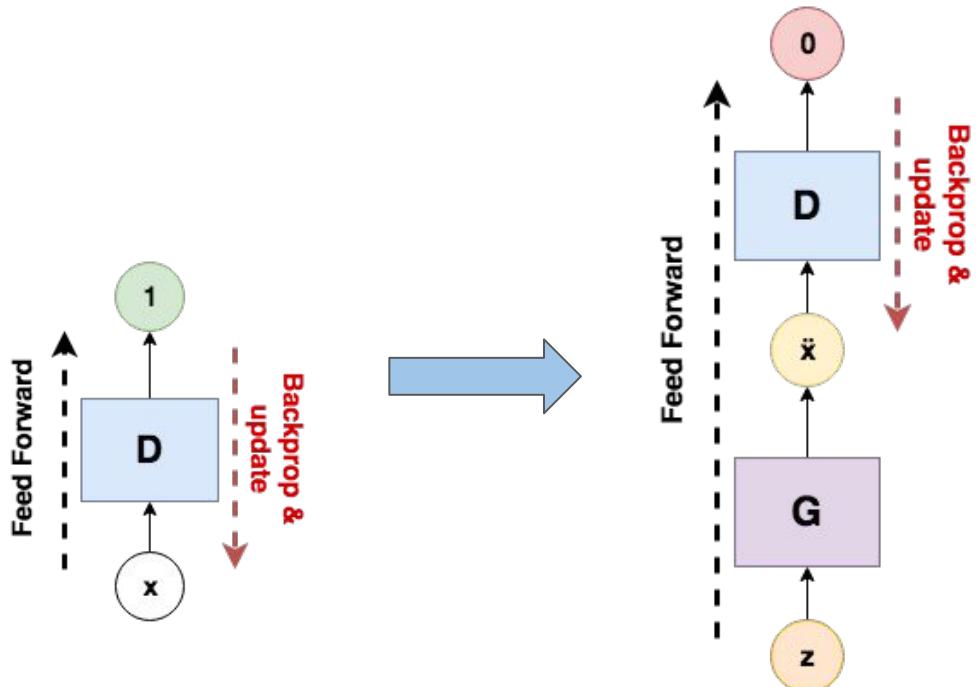
Adversarial training in a nutshell

- Pick a real scanpath x from training set
- Show x to D and update weights to predict “real” (1).



Adversarial training in a nutshell

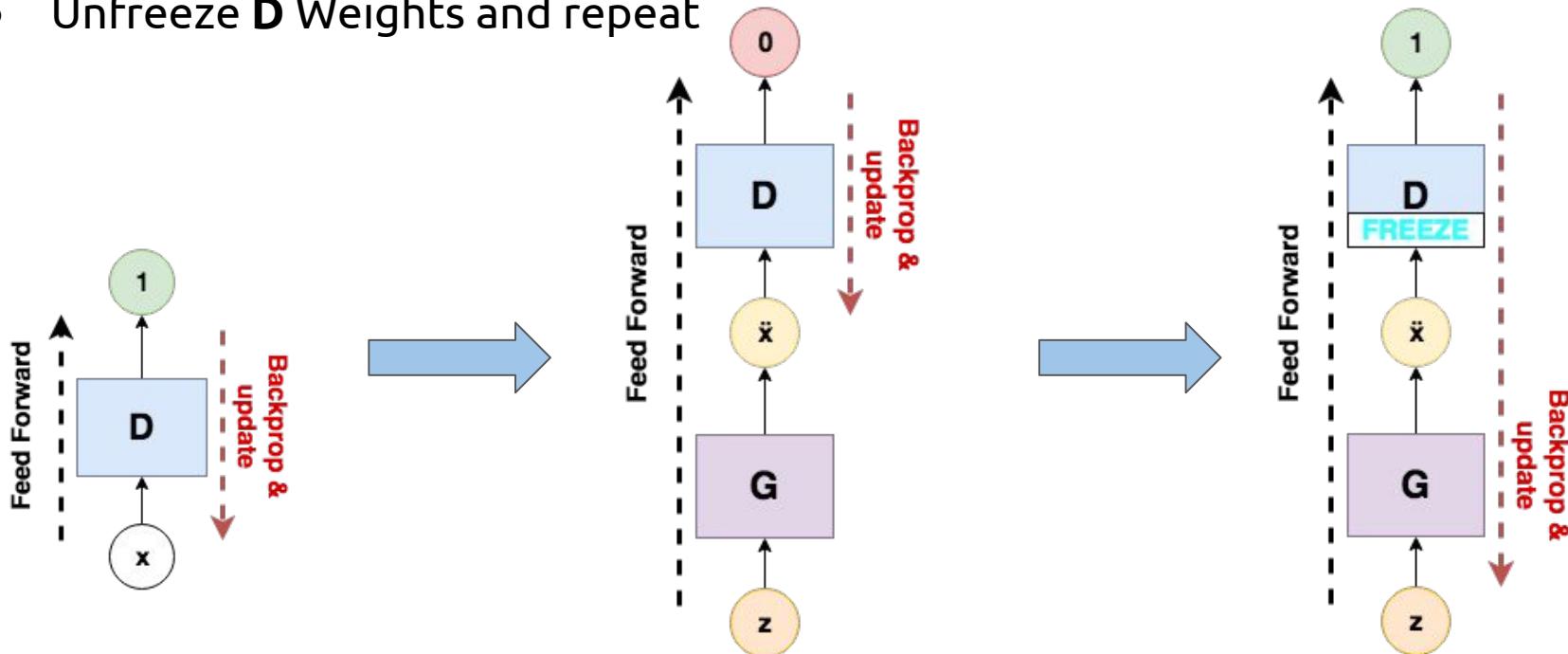
- **G** generates a synthetic scanpath \ddot{x} (adding noise & drop out)
- Show synthetic scanpath \ddot{x} to **D** and update its weights to predict “**fake**





Adversarial training in a nutshell

- Freeze **D** weights
- Keep showing the same synthetic scanpath \hat{x} to D
- Update **G** weights to make **D** predict “real” (1) (only **G** weights!)
- Unfreeze **D** Weights and repeat





Multiobjective Loss for G

$$L = L_{\text{cGAN}}(G, D) + \alpha L_{L^2}(G)$$

Content Loss

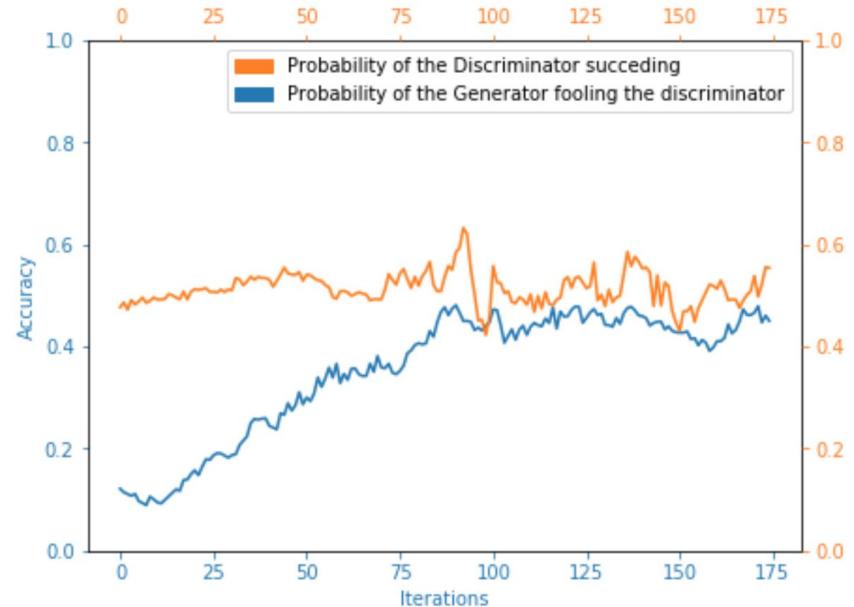
$$L_{L^2}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|^2]$$

Adversarial Loss for cGAN

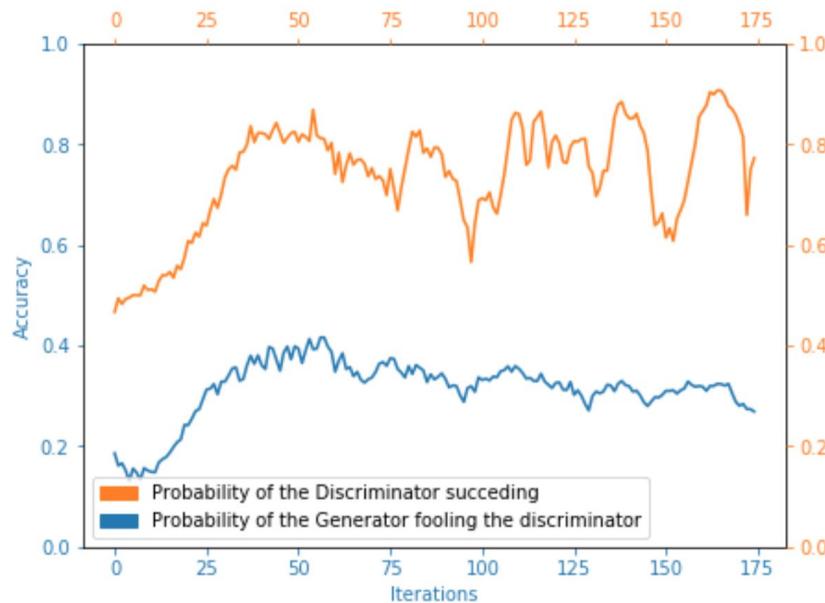
$$L_{\text{cGAN}}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]$$



Training curves



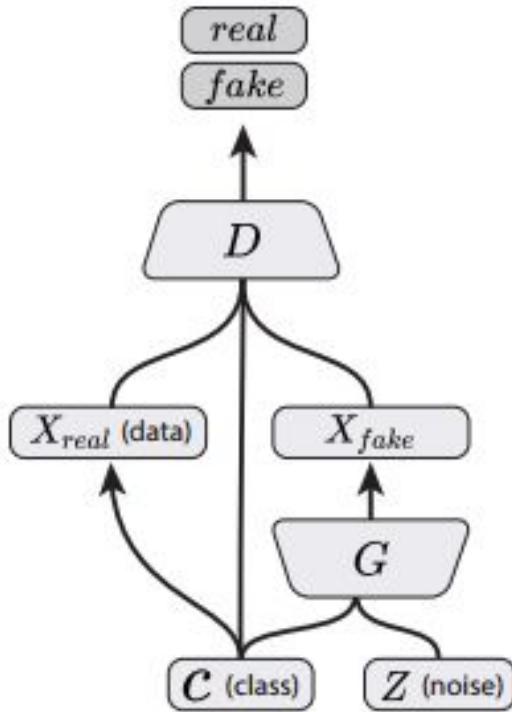
(a) Training with adversarial and content loss



(b) Training only with adversarial loss



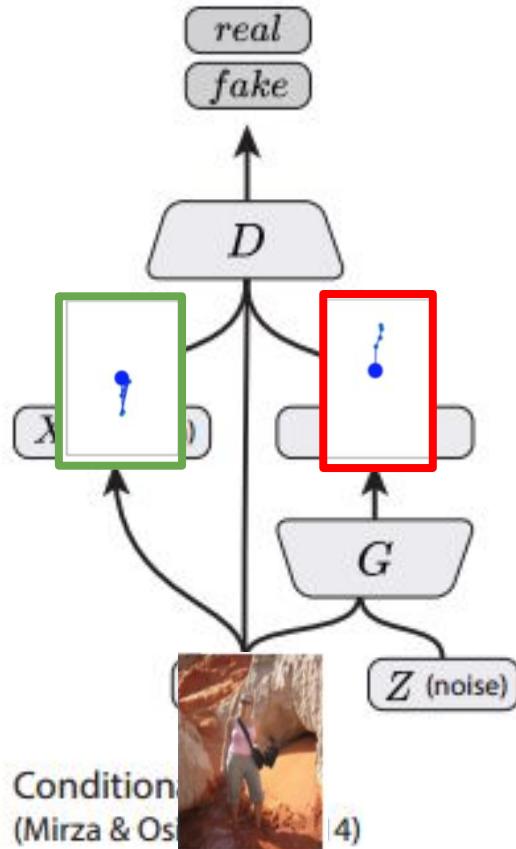
Model Architecture (cGAN)



Mirza, Mehdi, and Simon Osindero. "[Conditional generative adversarial nets.](#)" arXiv preprint arXiv:1411.1784 (2014).

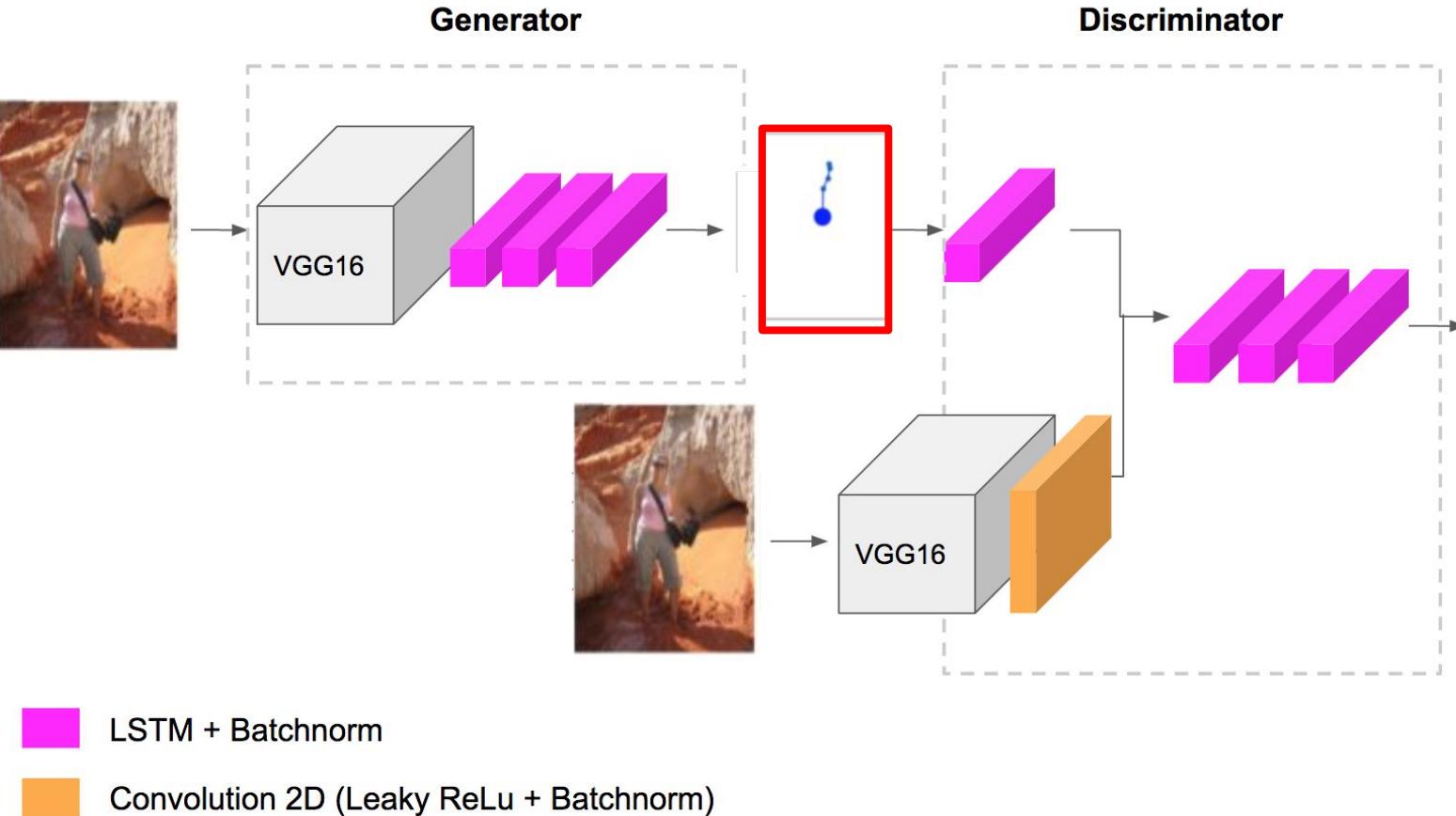


Model Architecture (cGAN)



Condition
(Mirza & Osipov 2014)

Model Architecture (cGAN)

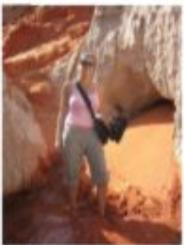


Qualitative Results: iSUN

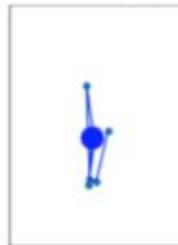
bit.ly/PathGAN
@DocXavi



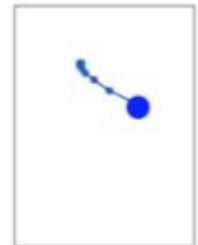
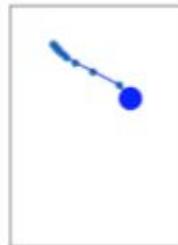
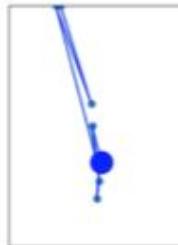
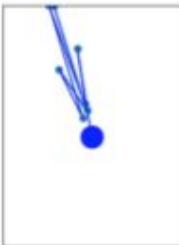
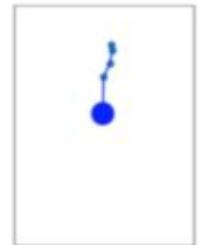
Stimuli



Ground Truth



Predictions





Quantitative Results: iSUN

Jarodzka ↓

SalTiNet	0.69
PathGAN without content loss	0.42
SalTiNet (fine-tuned on iSUN)	0.40
PathGAN	0.13

Qualitative Results: Salient360₍₂₀₁₇₎

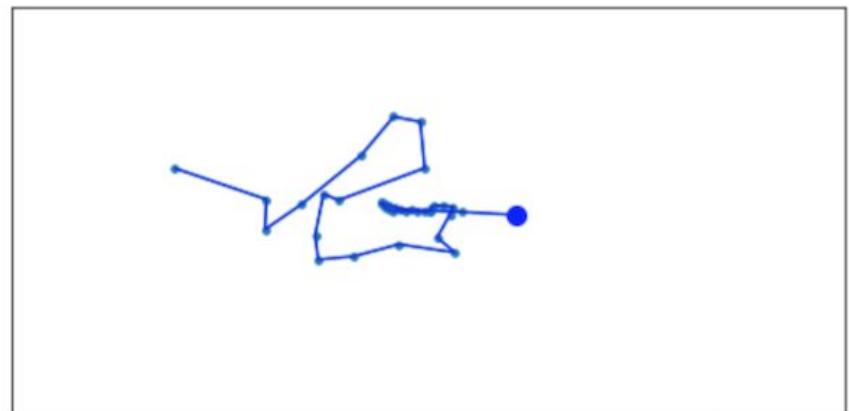
bit.ly/PathGAN
@DocXavi



Stimuli



Prediction



Quantitative Results: Salient360₍₂₀₁₇₎

bit.ly/PathGAN
@DocXavi



	Wuhan University	SJTU	SaltiNet	PathGAN
Jarodzka ↓	5.9517	4.6565	2.8697	0.74

2nd PLACE GRAND CHALLENGES

IEEE
ICME 2018
International Conference
on Multimedia and Expo



This Certificate Is Presented To

Marc Assens (Insight Centre for Data Analytics - Dublin City University),

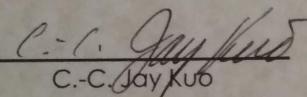
Xavier Giro-i-Nieto (Universitat Politècnica de Catalunya),

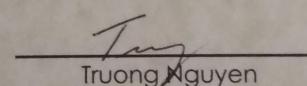
Kevin Mc Guinness (Insight Centre for Data Analytics - Dublin City University)

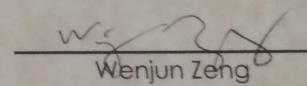
& Noel E. O'Connor (Insight Centre for Data Analytics - Dublin City University)

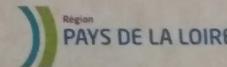
In recognition of being selected for the ICME 2018 Grand Challenge on Salient360° –
Visual Attention Modeling For 360° Content for the tracks of Prediction of
Eye-gaze Scan-paths for Images and Head-gaze Scan-paths for Images.

ICME 2018 General Chairs:


C.-C. Jay Kuo


Truong Nguyen

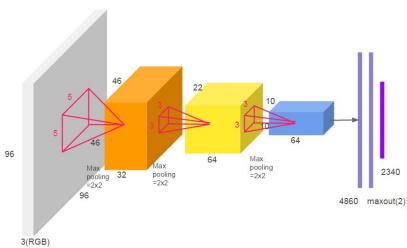

Wenjun Zeng



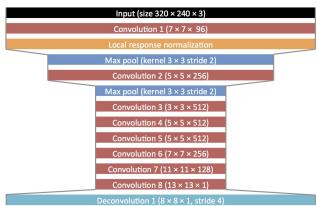
Saliency maps



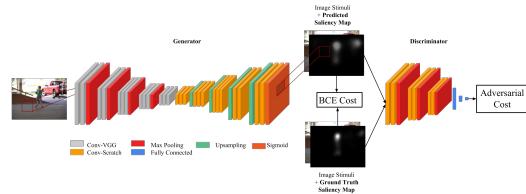
JuntingNet



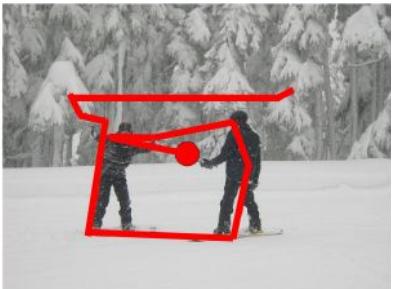
SalNet



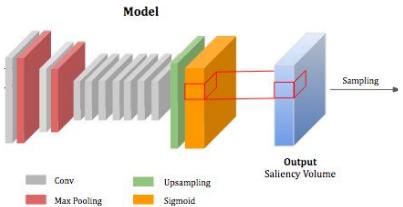
SalGAN



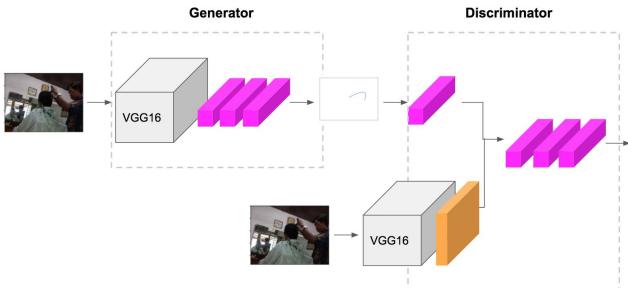
Scanpaths



SaltiNet



PathGAN

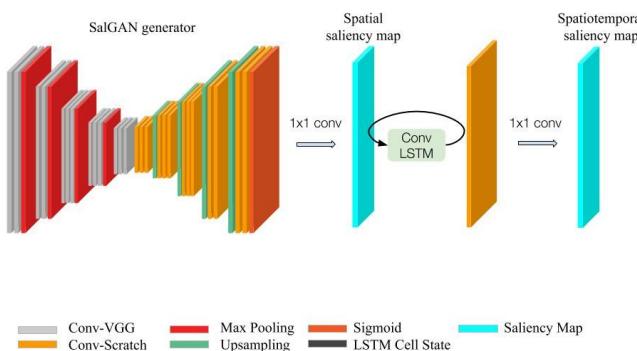


Next: What about video ?

bit.ly/PathGAN
@DocXavi



SalGAN + ConvLSTM



EgoMon dataset



Saliency maps



Check our other poster presented by Eva Mohedano !

Panagiotis Linardos, **Eva Mohedano**, Monica Cherto, Cathal Gurrin, Xavier Giro-i-Nieto. "[Temporal Saliency Adaptation in Egocentric Videos](#)", Extended abstract at the ECCV Workshop on Egocentric Perception, Interaction and Computing (EPIC), 2018.

Project page at: <http://bit.ly/pathgan>

Danke
schön !

AUTHORS PUBLICATION MODEL CODE EXAMPLES ACKNOWLEDGEMENTS

Fork me on GitHub

PathGAN: Visual Scanpath Prediction with Generative Adversarial Networks



Marc Assens Kevin McGuinness Xavier Giro Noel O'Connor



Insight
Centre for Data Analytics
Universitat Politècnica de Catalunya
BarceloNATech

Insight Center for Data Analytics (DCU)
Universitat Politècnica de Catalunya



GitHub

 PyTorch



xavier.giro@upc.edu



Generalitat de Catalunya
Government of Catalonia



 NVIDIA®