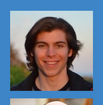
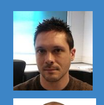
PathGAN: Visual Scanpath Prediction with Generative Adversarial Networks

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Barcelona Supercomputing Center Centro Nacional de Supercomputación


Marc Assens


Noel E. O'Connor


Kevin McGuinness


Xavier Giro-i-Nieto

DCU Dublin City University

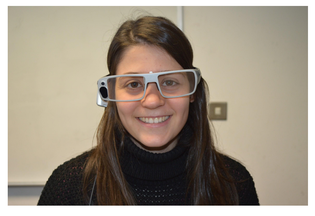
Insight

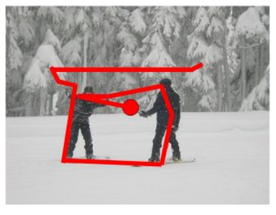
EPIC & 360PI Workshops @ ECCV18

ECCV 2018 European Conference on Computer Vision 8 – 14 September 2018 | Munich, Germany

Abstract

We introduce PathGAN, a deep neural network for visual scanpath prediction trained on adversarial examples. A visual scanpath is defined as the sequence of fixation points over an image defined by a human observer with its gaze. PathGAN is composed of two parts, the generator and the discriminator. Both parts extract features from images using off-the-shelf networks, and train recurrent layers to generate or discriminate scanpaths accordingly. In scanpath prediction, the stochastic nature of the data makes it very difficult to generate realistic predictions using supervised learning strategies, but we adopt **adversarial training** as a suitable alternative. Our experiments prove how PathGAN improves the state of the art of visual scanpath prediction on the iSUN and Salient360! datasets.



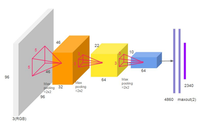


The UPC-DCU Saliency Zoo

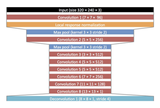
Saliency maps

Scanpaths

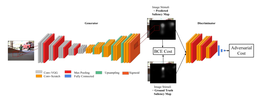
JuntingNet [1]



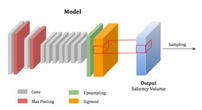
SalNet [1]



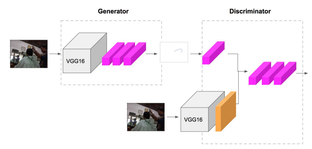
SalGAN [2]

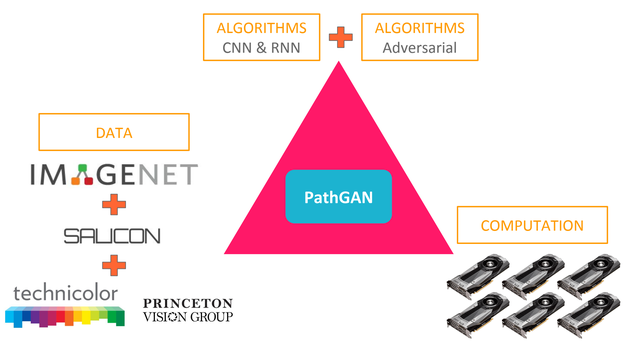


SaltiNet [3]

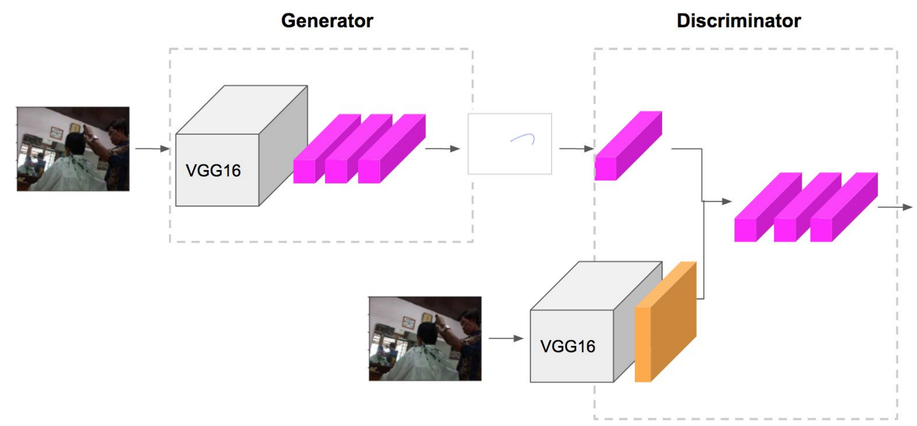


PathGAN





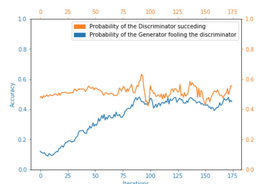
PathGAN



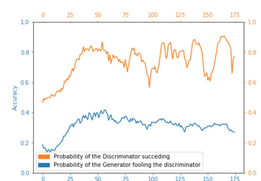
$$L = \underbrace{L_{\text{cGAN}}(G, D)}_{\text{GAN Loss}} + \underbrace{\alpha L_{L^2}(G)}_{\text{Content Loss}}$$

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

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



(a) Training with adversarial and content loss

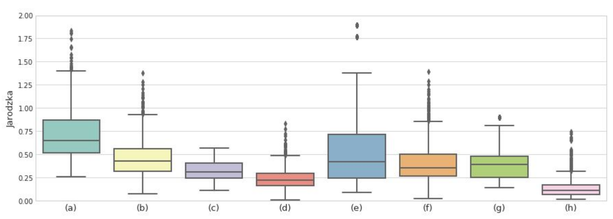


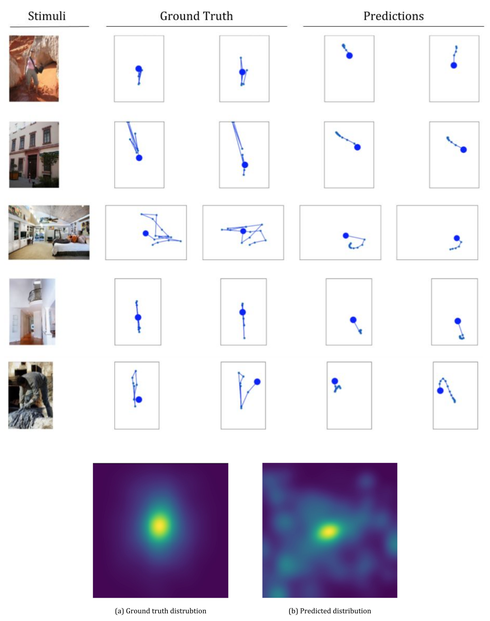
(b) Training only with adversarial loss

Results

Performance on iSUN dataset

id		Jarodzka↓
a	Random positions and number of fixations	0.71
b	Random positions and GT number of fixations	0.45
c	Sampling ground truth saliency maps	0.31
d	Interchanging scanpaths across images	0.23
e	SalTiNet	0.69
f	PathGAN without content loss	0.42
g	SalTiNet (fine-tuned on iSUN)	0.40
h	**PathGAN**	**0.13**



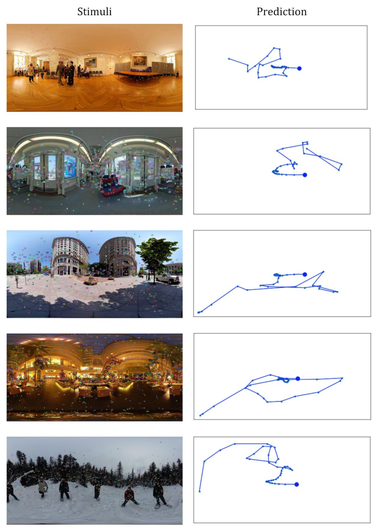


(a) Ground truth distrubtion

(b) Predicted distribution

Performance on the ICME Salient360! 2017 dataset

	Wuhan University	SJTU	SaltiNet	**PathGAN**
Jarodzka ↓	5.9517	4.6565	2.8697	**0.74**



2nd Place @ ICME Salient360! 2018 challenge

PyTorch

Model and source code: http://bit.ly/pathgan

References:
[1] Junting Pan et al. ""Shallow and deep convolutional networks for saliency prediction." CVPR 2016.
[2] Junting Pan et al. "SalGAN: Visual saliency prediction with generative adversarial networks." arXiv preprint (2017).
[3] Marc Assens et al. "SaltiNet:" ICCV 2017 EPIC Workshop.

Science Foundation Ireland sfi For what's next

Generalitat de Catalunya

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD

NVIDIA