SalGAN: visual saliency prediction with adversarial networks

이상용 / 2020-03-27

Computational Data Science LAB

SalGAN: visual saliency prediction with adversarial networks

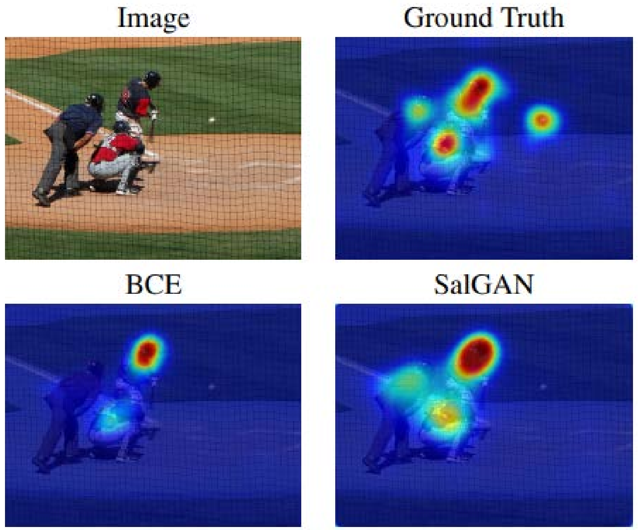
Computational Data Science LAB	
목차	1. Introduction 2. Architecture 3. Training 4. Experiments
논의사항 및 결정사항	
관련문서	Pan, J., Sayrol, E., Nieto, X. G. I., Ferrer, C. C., Torres, J., McGuinness, K., & OConnor, N. E. (2017, July). Salgan: Visual saliency prediction with adversarial networks. In CVPR Scene Understanding Workshop (SUNw).

CONTENTS

1. Introduction
2. Architecture
3. Training
4. Experiments

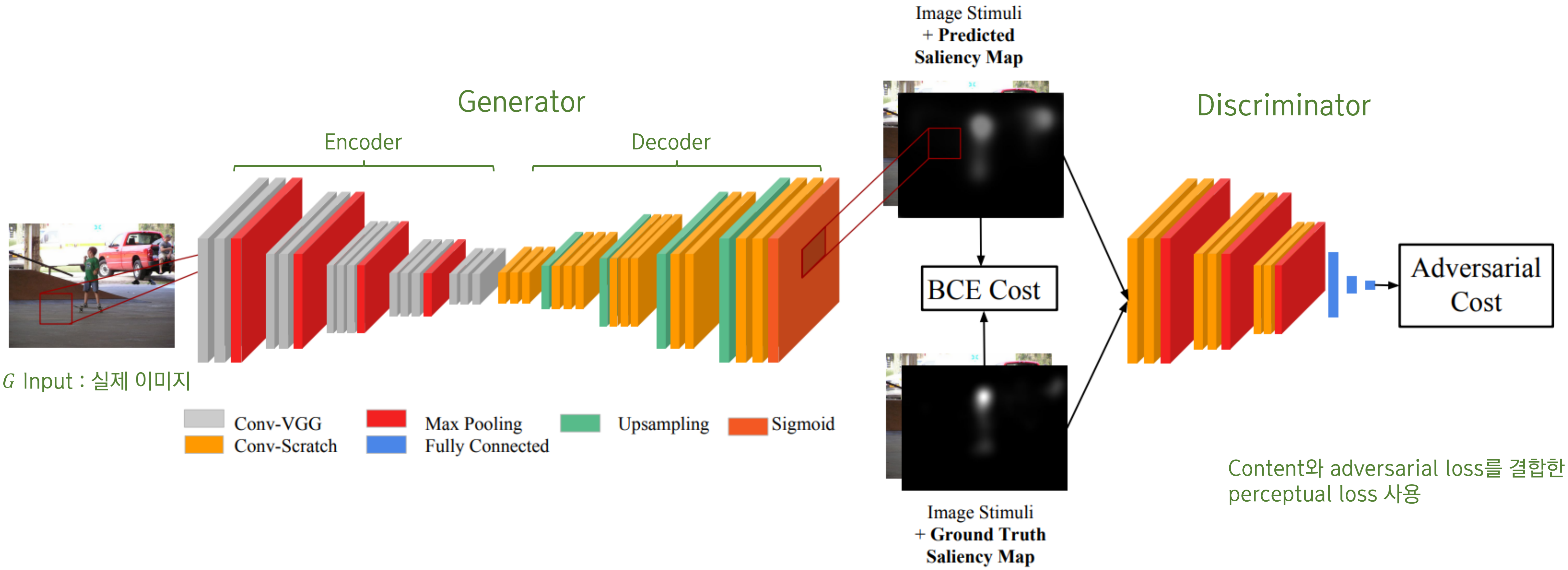
01 Introduction

- 최근의 saliency prediction 접근법은 일반적으로 single saliency metric기반의 loss function으로 학습
- 본 논문은 data-driven metric 기반의 saliency prediction method를 제안
- GAN 기반의 SalGAN (Saliency GAN)을 제안하며, generator와 discriminator의 adversarial loss function을 사용



SALICON(SALIENCY IN CONTEXT)
MIT saliency benchmark
…

- SalGAN



G Input : 실제 이미지

Content와 adversarial loss를 결합한 perceptual loss 사용

G Output : saliency map

02 | Architecture

- Generator architecture

layer	depth	kernel	stride	pad	activation
conv1_1	64	1×1	1	1	ReLU
conv1_2	64	3×3	1	1	ReLU
pool1		2×2	2	0	-
conv2_1	128	3×3	1	1	ReLU
conv2_2	128	3×3	1	1	ReLU
pool2		2×2	2	0	-
conv3_1	256	3×3	1	1	ReLU
conv3_2	256	3×3	1	1	ReLU
conv3_3	256	3×3	1	1	ReLU

Encoder

⋮

upsample8		2×2	2	0	-
conv9_1	128	3×3	1	1	ReLU
conv9_2	128	3×3	1	1	ReLU
upsample9		2×2	2	0	-
conv10_1	64	3×3	1	1	ReLU
conv10_2	64	3×3	1	1	ReLU
output	1	1x1	1	0	Sigmoid

Decoder

*Ground Trurt*의 값 : [0,1]

- Discriminator architecture

layer	depth	kernel	stride	pad	activation
conv1_1	3	1×1	1	1	ReLU
conv1_2	32	3×3	1	1	ReLU
pool1		2×2	2	0	-
conv2_1	64	3×3	1	1	ReLU
conv2_2	64	3×3	1	1	ReLU
pool2		2×2	2	0	-
conv3_1	64	3×3	1	1	ReLU
conv3_2	64	3×3	1	1	ReLU
pool3		2×2	2	0	-
fc4	100	-	-	-	tanh
fc5	2	-	-	-	tanh
fc6	1	-	-	-	sigmoid

- *Content loss*는 픽셀단위를 기준으로 각 saliency map이 ground truth와 얼마나 유사한지 측정
  - ✓ $N = W \times N$ 차원의 이미지 I가 주어졌을 때, saliency map S는 확률 벡터이며, S_j는 j번째 픽셀의 확률

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{j=1}^{N}\left(S_j - \hat{S}_j\right)^2$$

  - ✓ MSE metric은 baseline으로 사용

- 실험에서는 binary cross entropy(BCE) 사용
  - ✓ MSE는 saliency map에서 blurred contours를 주는 특성으로 인해 각 픽셀들을 독립적으로 취급하는 BCE 이용

03 | Training

- *Content loss (BCE)*

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{j=1}^{N}\left(S_j \log\left(\hat{S}_j\right) + \left(1 - S_j\right)\log\left(1 - \hat{S}_j\right)\right)$$

- *Adversarial loss*
  - ✓ Content loss + discriminator loss

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{BCE} + L\left(D\left(I, \hat{S}\right), 1\right)$$

Content loss + discriminator loss

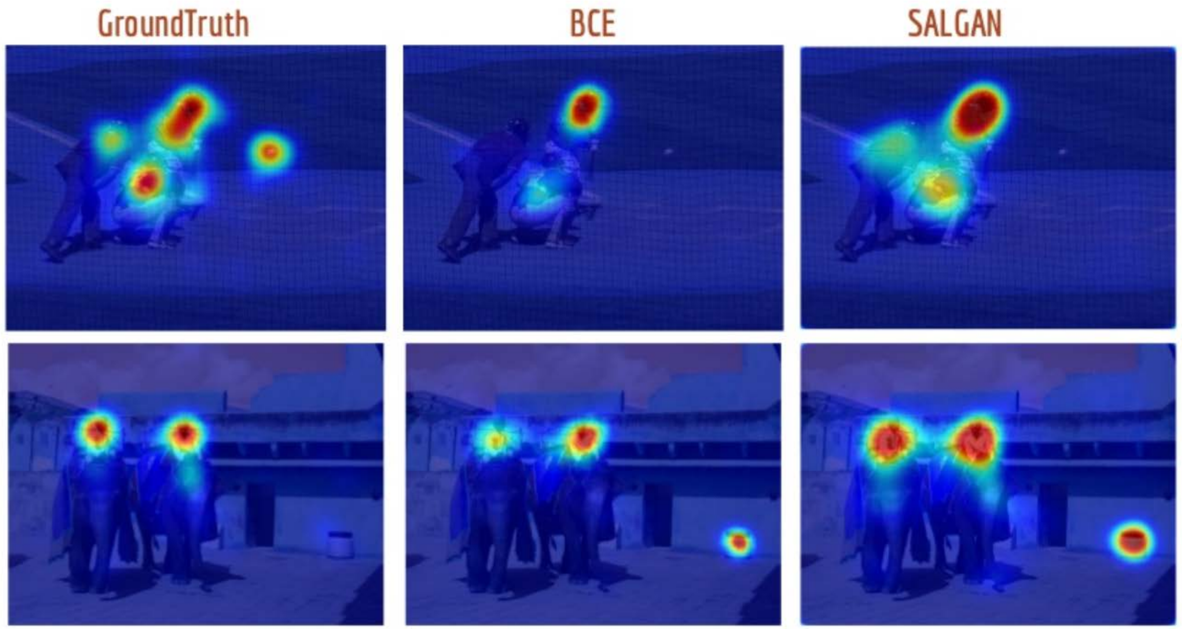
- How SalGAN is different from traditional GANs?

1. 전통적인 GAN은 입력으로 **random noise**가 들어가서 realistic **images**를 출력하지만, SalGAN은 입력으로 **실제 이미지**가 들어가고 realistic **saliency map**이 출력으로 반환 됨

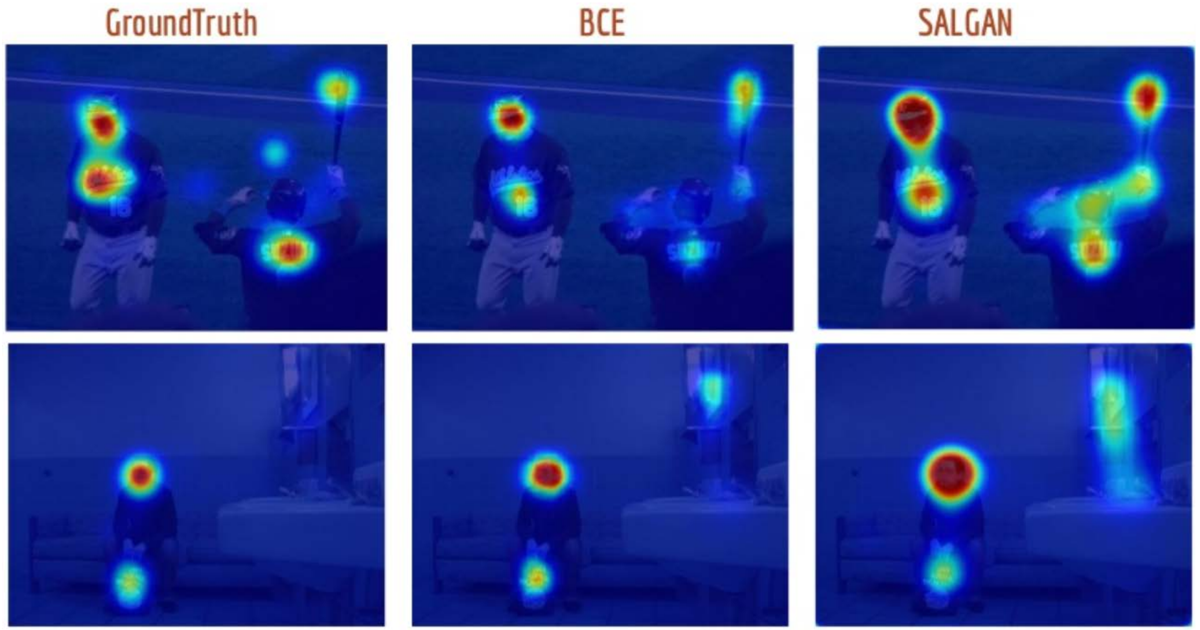
2. 전통적인 GAN은 Discriminator의 입력으로 real · fake 이미지가 들어가지만, SalGAN은 무조건 input image에 대응되는 realistic saliency map을 만들기 때문에, discriminator의 입력으로 **실제 이미지**와 **생성된 saliency map** pair가 들어감

3. GAN은 생성된 이미지에 대응하는 ground truth가 없지만, SalGAN은 ground truth saliency map이 존재

- Qualitative Results



- Qualitative Results – Failure case



- Quantitative Results

SALICON (test)	AUC-J ↑	Sim ↑	EMD ↓	AUC-B ↑	sAUC ↑	CC ↑	NSS ↑	KL ↓
DSCLRCN [24](*)	-	-	-	0.884	0.776	0.831	3.157	-
SalGAN	-	-	-	**0.884**	**0.772**	**0.781**	**2.459**	-
ML-NET [5]	-	-	-	(0.866)	(0.768)	(0.743)	2.789	-
SalNet [25]	-	-	-	(0.858)	(0.724)	(0.609)	(1.859)	-
MIT300	**AUC-J ↑**	**Sim ↑**	**EMD ↓**	**AUC-B ↑**	**sAUC ↑**	**CC ↑**	**NSS ↑**	**KL ↓**
Humans	0.92	1.00	0.00	0.88	0.81	1.0	3.29	0.00
Deep Gaze II [21](*)	0.88	(0.46)	(3.98)	0.86	0.72	(0.52)	(1.29)	(0.96)
DSCLRCN [24](*)	0.87	0.68	2.17	(0.79)	0.72	0.80	2.35	0.95
DeepFix [17](*)	0.87	0.67	2.04	(0.80)	(0.71)	0.78	2.26	0.63
SALICON [9]	0.87	(0.60)	(2.62)	0.85	0.74	0.74	2.12	0.54
SalGAN	**0.86**	**0.63**	**2.29**	**0.81**	**0.72**	**0.73**	**2.04**	**1.07**
PDP [11]	(0.85)	(0.60)	(2.58)	(0.80)	0.73	(0.70)	2.05	0.92
ML-NET [5]	(0.85)	(0.59)	(2.63)	(0.75)	(0.70)	(0.67)	2.05	(1.10)
Deep Gaze I [19]	(0.84)	(0.39)	(4.97)	0.83	(0.66)	(0.48)	(1.22)	(1.23)
iSEEL [29](*)	(0.84)	(0.57)	(2.72)	0.81	(0.68)	(0.65)	(1.78)	0.65
SalNet [25]	(0.83)	(0.52)	(3.31)	0.82	(0.69)	(0.58)	(1.51)	0.81
BMS [31]	(0.83)	(0.51)	(3.35)	0.82	(0.65)	(0.55)	(1.41)	0.81

Q&A

감사합니다.