Neural networks in image saliency prediction

...

Magdalena Rajtor

Overview of image saliency

❖ saliency - unique features of a processed image
saliency map - their topographical representation
❖ more than just object detection
❖ modelling approaches:
  ➢ Graph-Based Visual Saliency (GBVS)
  ➢ Contextual Guidance Model

  ➢ CNNs, RNNs, GANs
  ➢ CNN+ (e.g. SAM-ResNet)
  ➢ CNN++ (e.g. TranSalNet)

Lou, J., Lin, H., Marshall, D., Saupe, D. & Liu, H. (2022). TranSalNet: Towards perceptually relevant visual saliency prediction, *Neurocomputing, 494*, 455-467, https://doi.org/10.1016/j.neucom.2022.04.080

Challenges with modelling saliency

- they should have semantics (understanding image content)
- task and context dependance
- emotional factors
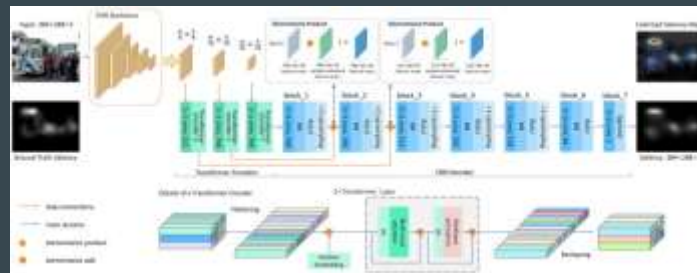
datasets: CAT2000 (sample of 400 images), EMOd (~700), paintings (150)
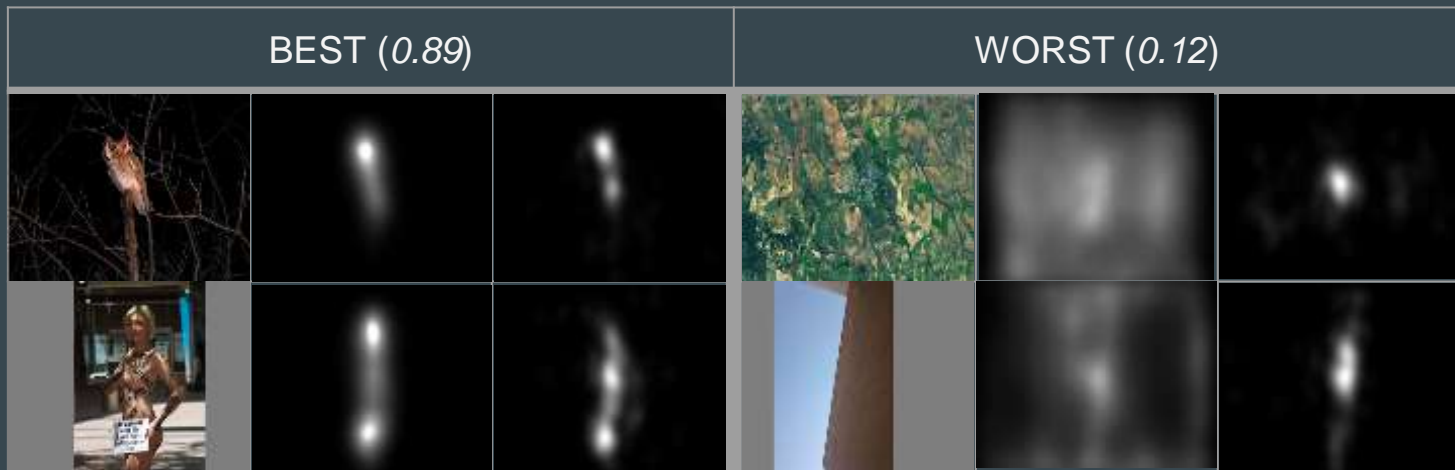
TranSalNet_Res model
SIM on CAT2000: *0.75*

SIM (Similarity Index) vs SSIM (Structural Similarity Index)
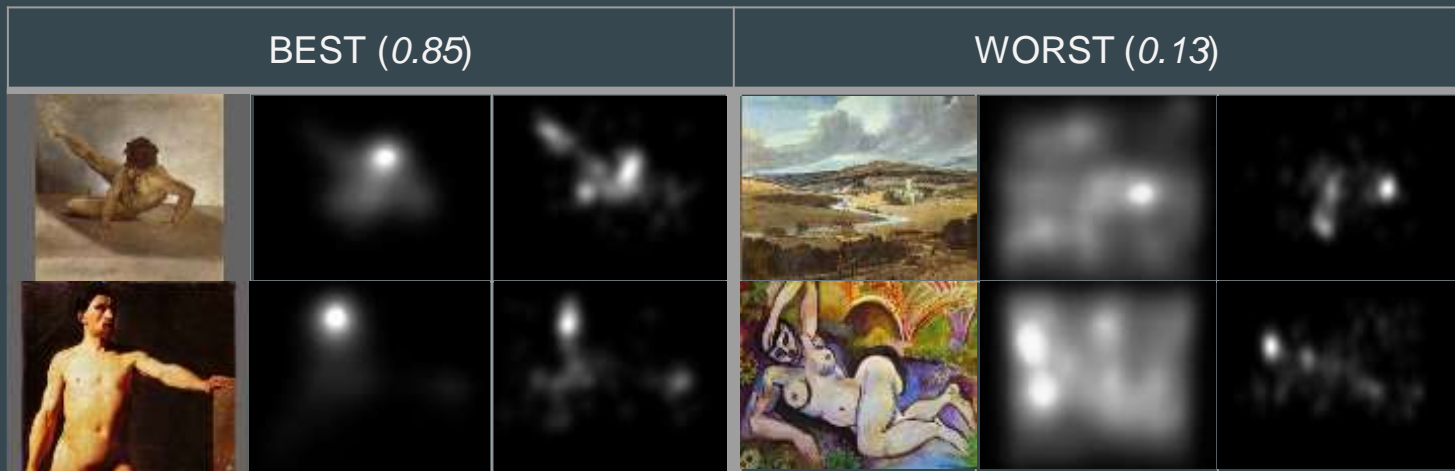
1) Baseline performance - CAT2000

CAT2000	*0.47*
ART	*0.48*
EMO	*0.57*

mean SSIM scores

BEST (*0.89*)	WORST (*0.12*)

1) Baseline performance - ART

CAT2000	*0.47*
ART	*0.48*
EMO	*0.57*

mean SSIM scores

1) Baseline performance - EMO

CAT2000	*0.47*
ART	*0.48*
EMO	*0.57*

mean SSIM scores

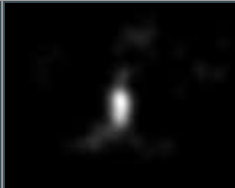| BEST (*0.87*)	WORST (*0.15*)
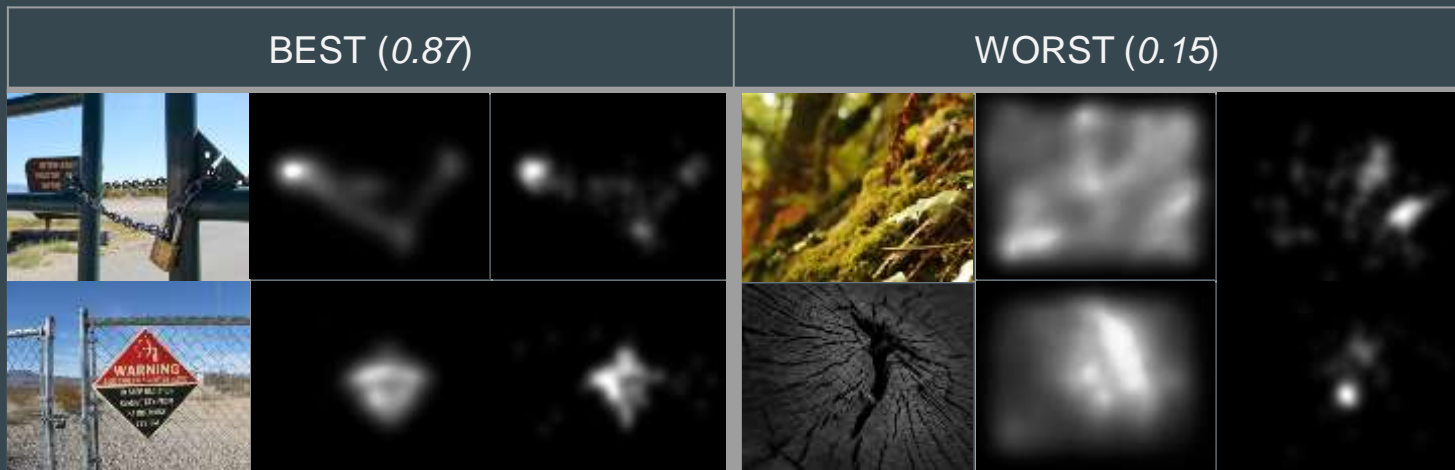
ART dataset

15%test

EMO dataset

15%test

70% train   15% val

70% train   15% val
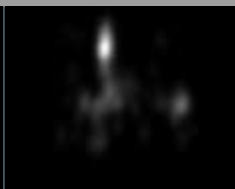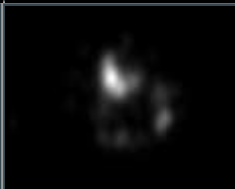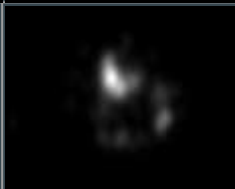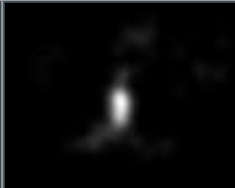
base model
model finetuned on

ART

2)   Finetuning on ART

		BEST			WORST		
ART base 0.49	best: _0.83_						
	worst: _0.13_						
ART 0.59	best: _0.77_						
	worst: _0.34_						
EMO base 0.57	best: _0.83_						
	worst: _0.31_						
EMO 0.59	best: _0.75_						
	worst: _0.41_						

3) Finetuning on EMO

		BEST			WORST		
ART base 0.49	best: _0.83_						
	worst: 0.13						
ART 0.59	best: 0.77						
	worst: _0.34_						
EMO base 0.57	best: 0.83						
	worst: 0.31						
EMO 0.68	best: _0.86_						
	worst: _0.41_						

Mean SSIM scores for each CAT2000 category