CA-Stream: Attention-based pooling for interpretable image recognition

The 3^{rd} Explainable AI Workshop for Computer Vision (XAI4CV) - CVPR 2024

Felipe Torres[1] Hanwei Zhang[2] Ronan Sicre[1]
Stéphane Ayache [1] Yannis Avrithis[3]

Centrale Marseille, Aix Marseille Univ, CNRS, LIS, France
Institute of Intelligent Software, China
Institute for Advanced Research in Artificial Intelligence (IARAI), Austria

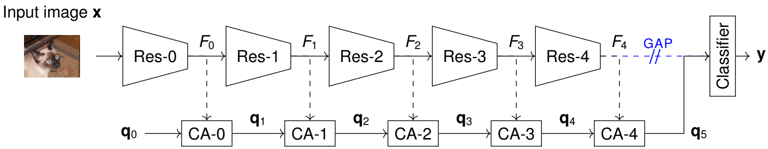
June 18^{th}, 2024

Main Idea

An attention based pooling to improve interpretability measurements.

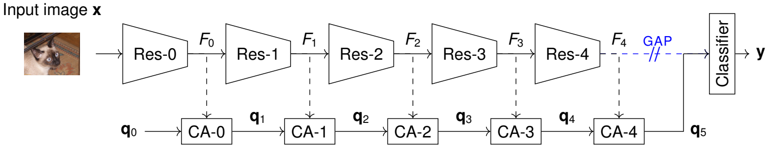
Main Idea

An attention based pooling to improve interpretability measurements.



Main Idea

An attention based pooling to improve interpretability measurements.

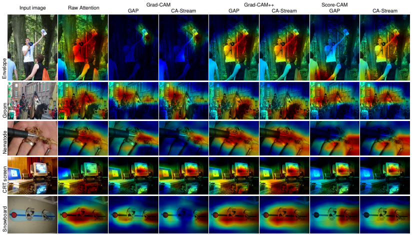


- Improve quantitative explanations for existing models.
- Provide class agnostic raw attention maps.
- Maintain recognition properties of each model.

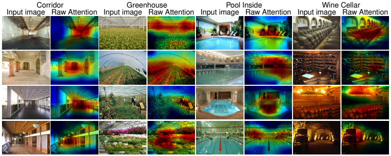
Our Results

Qualitative Results

CAM on ImageNet



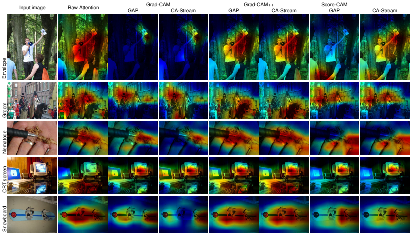
Raw Attention on MIT 67 dataset



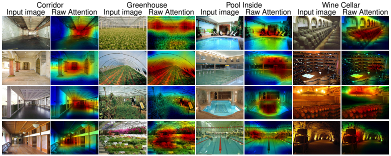
Our Results

Qualitative Results

CAM on ImageNet



Raw Attention on MIT 67 dataset



Quantitative Results

Interpretability results on ImageNet

NETWORK	POOLING	ACC↑
ResNet-50	GAP	74.55
	CA	74.70
ConvNeXt-B	GAP	83.72
	CA	83.51

NETWORK	ATTRIBUTION	POOLING	AD↓	AG↑	AI↑	I↑	D↓
RESNET-50	Grad-CAM	GAP	13.04	17.56	44.47	72.57	**13.24**
		CA	**12.54**	**22.67**	**48.56**	**75.53**	13.50
	Grad-CAM++	GAP	**13.79**	15.87	42.08	72.32	**13.33**
		CA	13.99	**19.29**	**44.60**	**75.21**	13.78
	Score-CAM	GAP	8.83	17.97	48.46	71.99	**14.31**
		CA	**7.09**	**23.65**	**54.20**	**74.91**	14.68
CONVNEXT-B	Grad-CAM	GAP	33.72	2.43	15.25	52.85	**29.57**
		CA	**19.45**	**13.96**	**32.89**	**86.38**	45.29
	Grad-CAM++	GAP	**34.01**	2.37	15.60	52.83	**29.17**
		CA	36.69	**8.00**	**21.95**	**85.39**	53.42
	Score-CAM	GAP	43.55	2.23	15.67	50.96	**39.49**
		CA	**23.51**	**11.04**	**27.35**	**83.41**	60.53