Image Attribution by Generating Images

Aniket Singh   Anoop Namboodiri

CVIT, IIIT Hyderabad, India

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY H Y D E R A B A D

Terminology
- **Attribution**: (a.k.a heat maps) Identifying where classifier looks to make its decision
- **CAM**: Class Activation Maps. A family of Attribution Methods (e.g. GradCAM, ScoreCAM)
- **Single Image Generation**: Making realistic variations of a single Image (e.g. SinGAN).

Motivation
- **(right)** If we can create variations of an image, we can see a broader range of the classifier's beaviour. This may help attribution.
- **(left)** t-sne of conv-featues from variations (red) and from original image (blue). red feats semantically interpolate between blue feats.
- But, need a fast generative model: SinGAN too slow (~hours/image),so we use Generative Patch Nearest Neighbors (GPNN) (~ seconds/image)
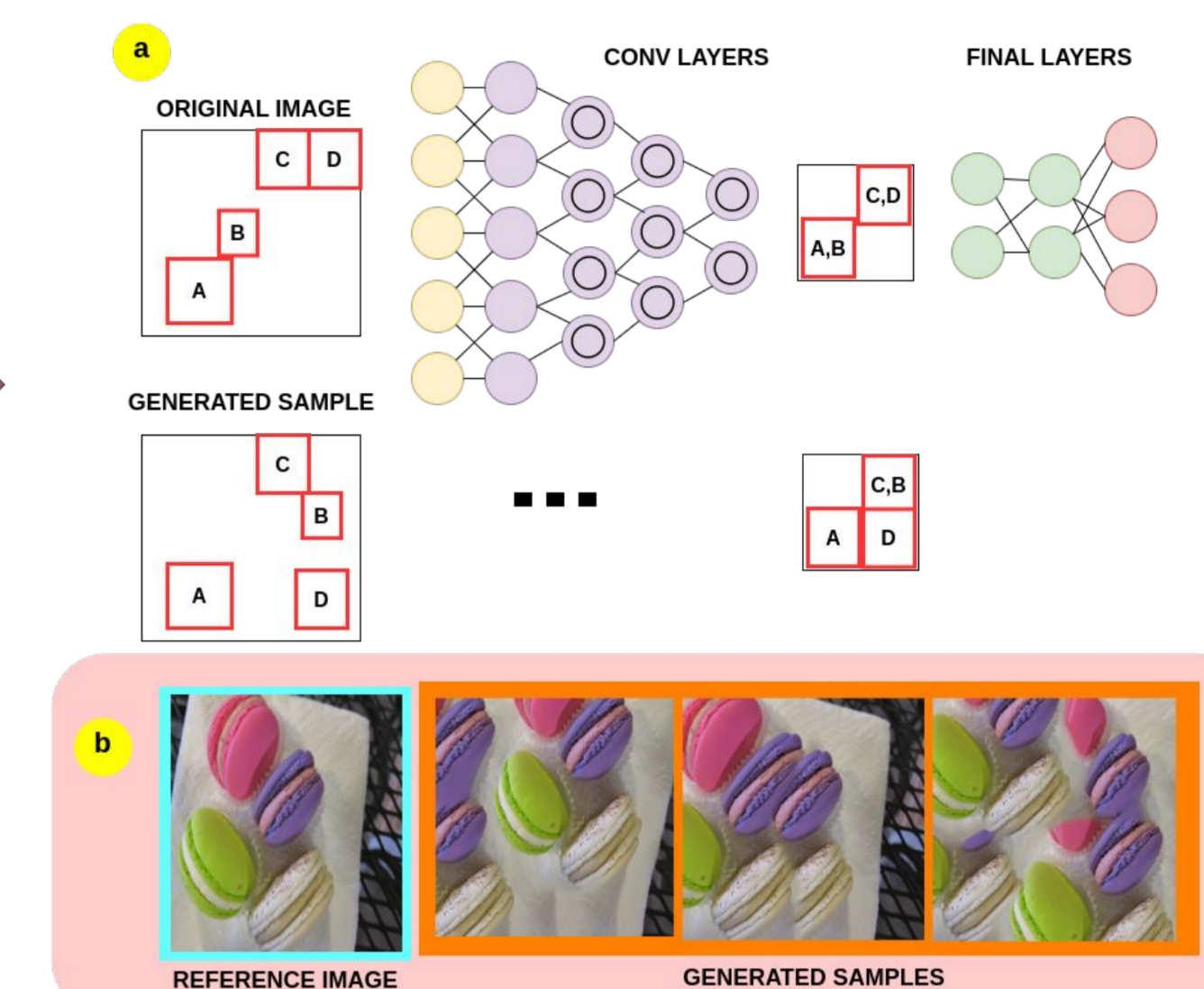

Contributions
- Investigating how generated variations of a image can help improve attribution.
- A new Attribution method, **GPNN-CAM** to effectively utilze these varations

Method 2 step pipeline **GPNN-CAM**
- **Phase 1(green)**: **GradCAM flavoured algorithm,** GradCAM: combine gradients (grads) and activations (acts) to get heat map. GPNN-CAM: Instead of single image, we use act and grad of all variations.
- **Phase 2 (yellow)**: Use lower conv-layer acts to regress to previous heat map.
- Combine heat maps from all layers
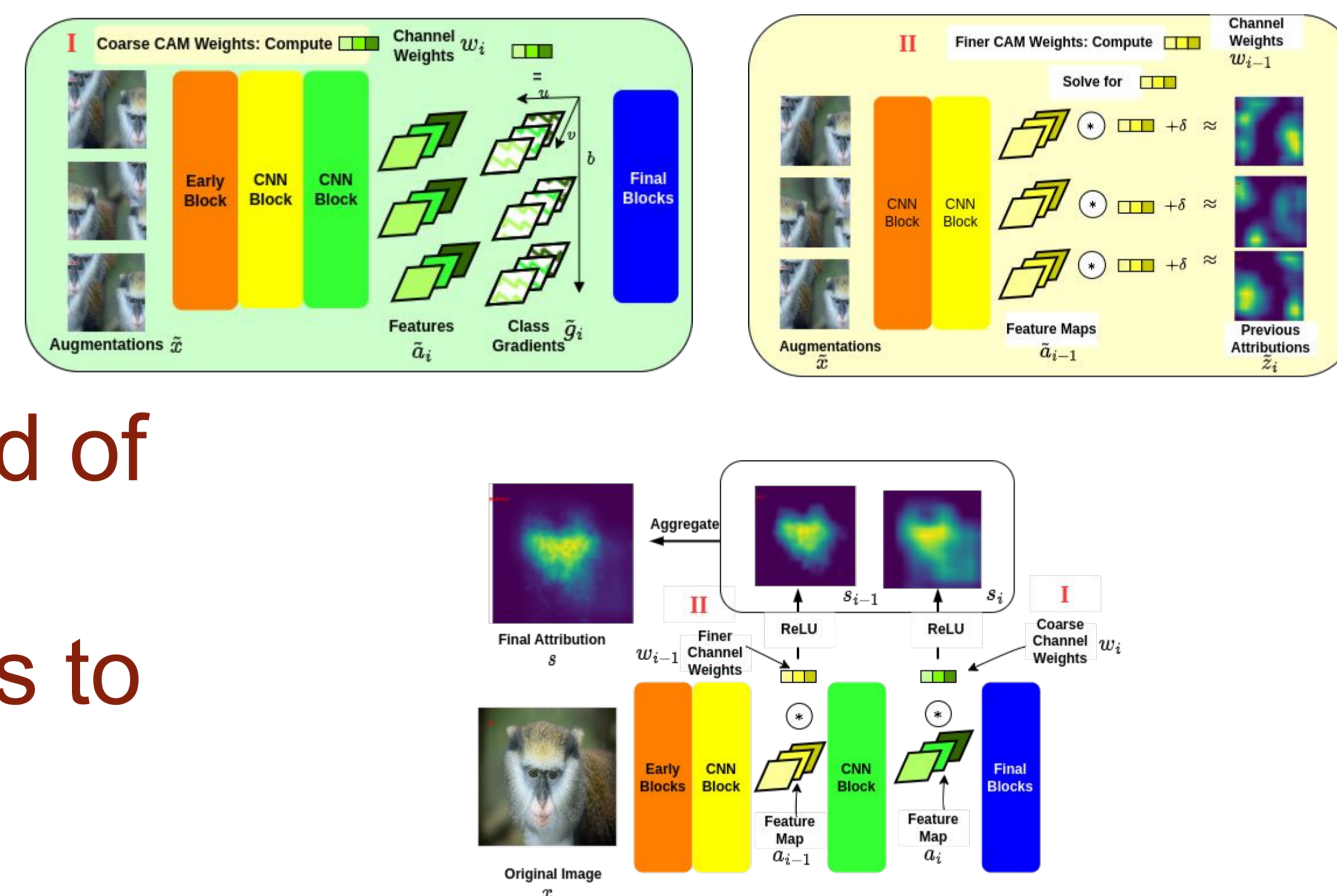
Qualitative Results
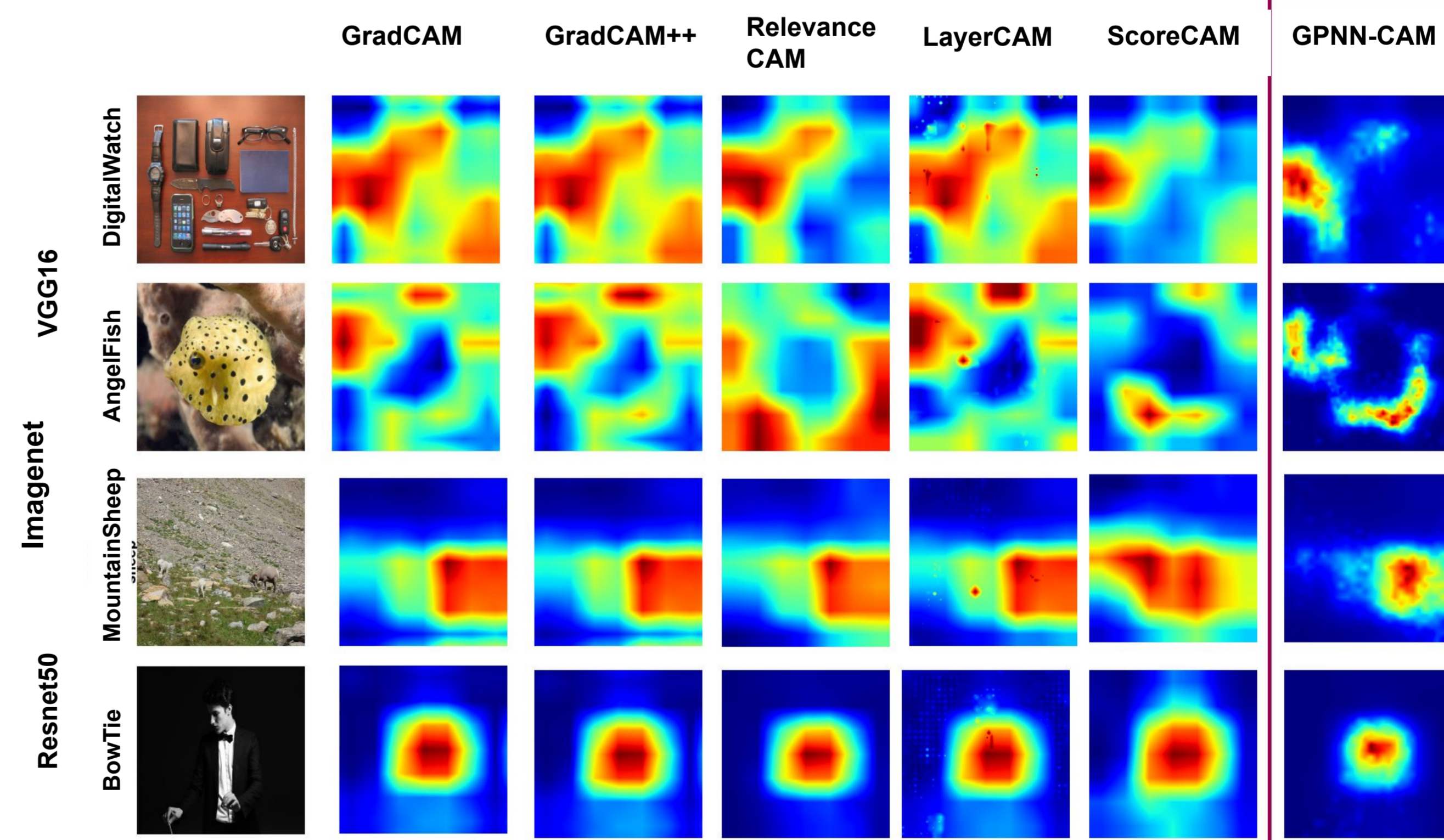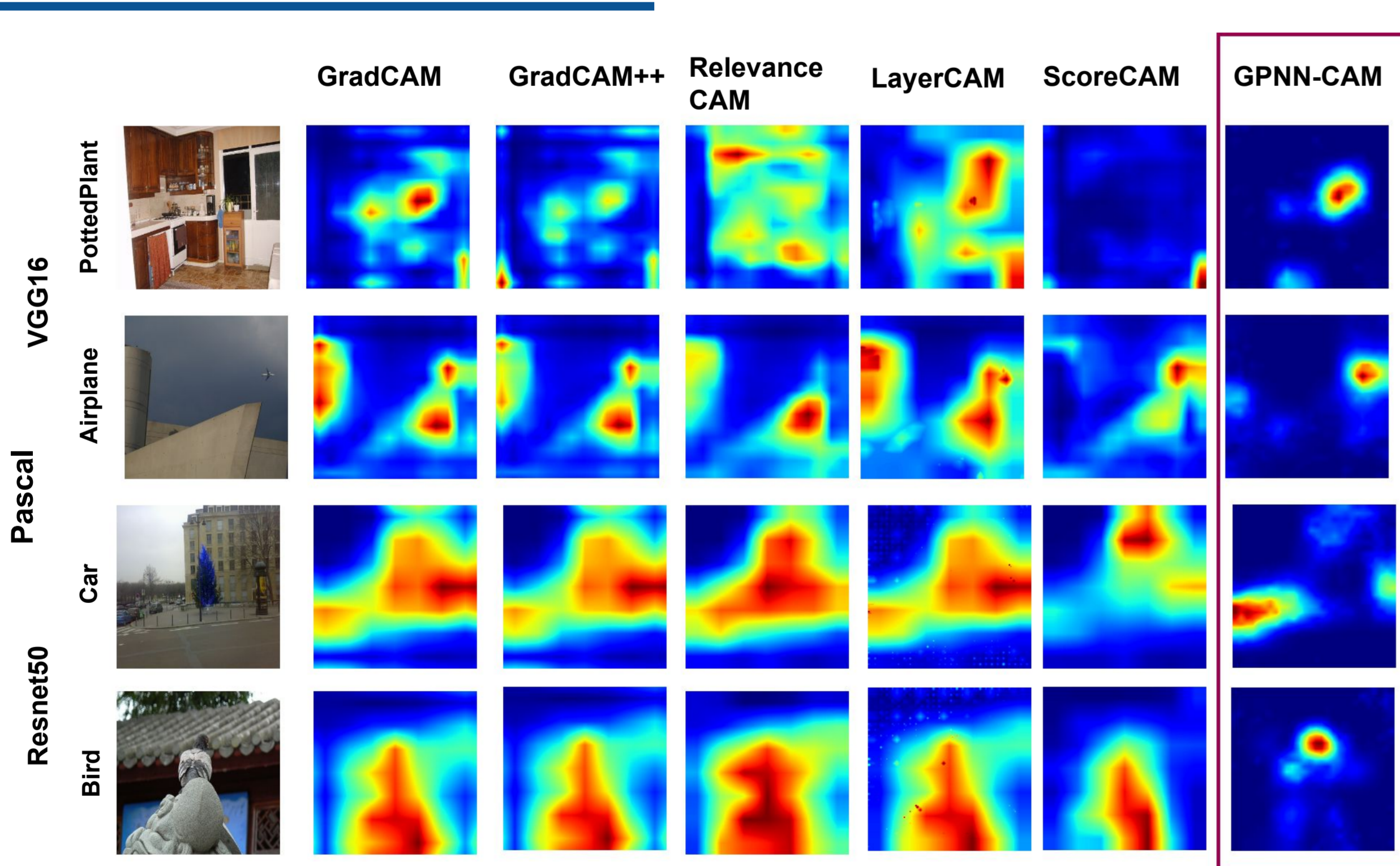
Quantitative Results
- <u>Pointing Game</u>: % of times the highest point of the heat map is inside the object bounding box. (Best in bold, second best underlined)

	Imagenet		Pascal(All/Diff)	
	VGG16	ResNet50	VGG16	ResNet50
GradCAM	<u>0.937</u>	0.928	<u>0.780</u>/<u>0.595</u>	0.765/0.552
GradCAM++	0.927	<u>0.929</u>	0.724/0.513	0.759/0.543
RelenvanceCAM	0.914	0.917	0.634/0.362	0.757/0.533
ScoreCAM	0.918	0.905	0.682/0.508	<u>0.789</u>/<u>0.624</u>
LayerCAM	0.899	0.917	0.750/0.557	0.771/0.611
GPNN-CAM	**0.946**	**0.945**	**0.830/0.699**	**0.866/0.749**

References
- Drop the gan: In defense of patches nearest neighbors as single image generative models, Granot et. al. CVPR 2022
- Grad-cam: Visual explanations from deep networks via gradient-based localization, Selvaraju et. al. ICCV 2017

Contact
- aniketsinghresearch@gmail.com