On Out-of-Distribution Detection Algorithms with Deep Neural Skin Cancer Classifiers

Andre G. C. Pacheco[1] Chandramouli S. Sastry[2, 3] Thomas Trappenberg[2]
Sageev Oore[2,3] Renato A. Krohling[1]

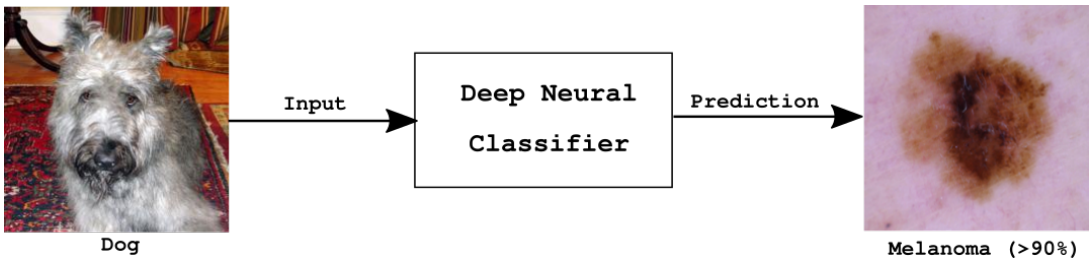
[1]Federal University of Espirito Santo - Vitória, Brazil
[2] Dalhousie University - Halifax, Canada
[3] Vector Institute - Toronto, Canada

{agcpacheco, rkrohling}@inf.ufes.br, cssastry@dal.ca, {tt,sageev}@cs.dal.ca

INTRODUCTION

What are out-of-distribution (OOD) samples?

- Samples that do not contain any of the labels modeled during training phase



INTRODUCTION

Problem:

- Deep Neural Softmax classifiers make over-confident predictions for OOD samples
- Detecting OOD samples is challenging

Objective:

- Detecting such OOD samples, in particular for skin cancer classification

INTRODUCTION

We examine the performance of the OOD detection algorithms with skin cancer classifiers

- State-of-the-art OOD algorithms:
  - ODIN (Liang et al., 2017)
  - Mahalanobis (Lee et al., 2018)
  - Gram-OOD (Sastry and Oore, 2019)

- Gram-OOD*:
  - An extension of the Gram-OOD algorithm that generally performs better for this particular task

SUMMARY OF OOD ALGORITHMS

ODIN:

- Use softmax with temperature as confidence on perturbed inputs.
- Needs to fine-tune temperature and perturbation magnitude.

Mahalanobis:

- Computes layerwise Mahalanobis distances from class-conditional feature distributions.
- Mahalanobis distances are used to train a Logistic Regression Detector.
- Needs OOD samples to train the Logistic Regression Detector.

GRAM MATRIX OOD DETECTION

- Take into account intermediate feature activations
- Compute Gram Matrices at every layer and check for anomalously high or low values.
- Does not require any knowledge of OOD samples.
- Can work with any pre-trained model.

GRAM MATRIX

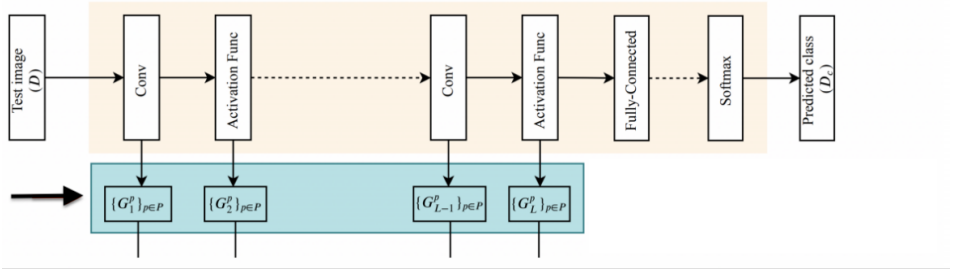
- Let F_l refer to the activations at layer l of shape $[C_l, H_l * W_l]$.
- Gram Matrix is computed using F_l as:

$$G_l = F_l F_l^\top \tag{1}$$

- Gram Matrix of Order p is computed as:

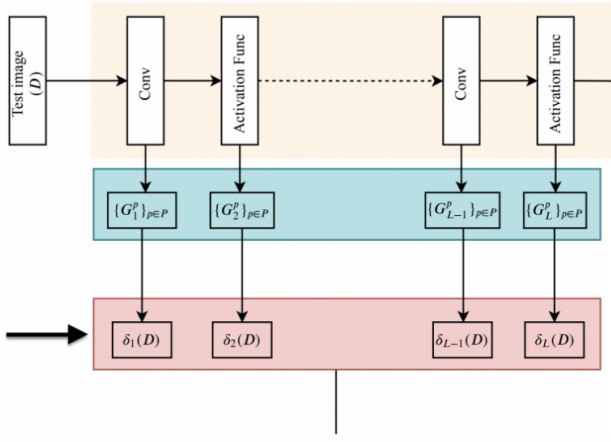
$$G_l^p = F_l^p F_l^{p\top} \tag{2}$$

GRAM MATRIX AS PAIRWISE CORRELATIONS



- Pairwise correlations between feature maps are computed using G_l^p of various orders

LAYERWISE DEVIATION

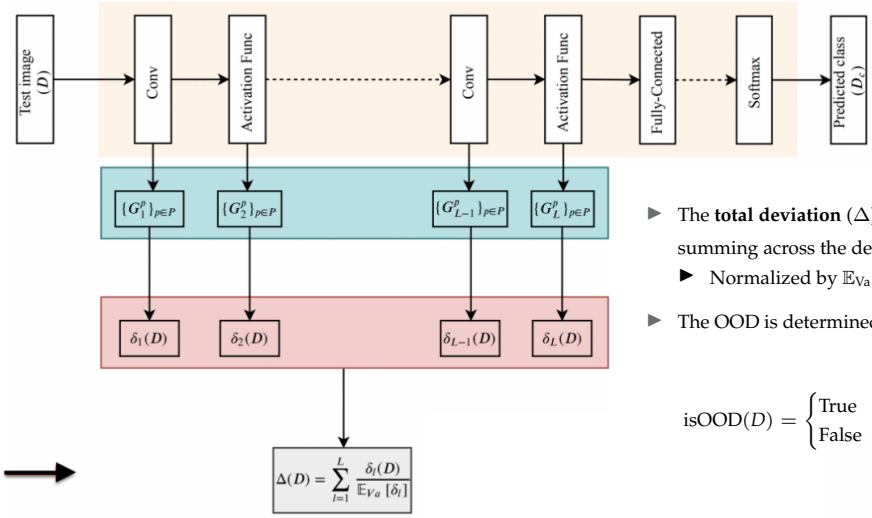


- **Lawerwise deviations** $\delta(D)$ are computed from the min and max of G_l^p w.r.t. the class:

$$\delta_l(\lambda_l, \Lambda_l, g_l) = \begin{cases} 0 & \text{if } \lambda_l \leq g_l \leq \Lambda_l \\ \frac{\lambda_l - g_l}{|\lambda_l|} & \text{if } g_l < \lambda_l \\ \frac{g_l - \Lambda_l}{|\Lambda_l|} & \text{if } g_l > \Lambda_l \end{cases}$$

where $\lambda_l = \min\left[G_l^p\right]$ and $\Lambda_l = \max\left[G_l^p\right]$

TOTAL DEVIATION



- The **total deviation** (Δ) is computing by summing across the deviations of all layers
  - Normalized by $\mathbb{E}_{\text{Va}}[\delta_l]$
- The OOD is determined as follows:

$$\text{isOOD}(D) = \begin{cases} \text{True} & \text{if } \Delta(D) > \tau \\ \text{False} & \text{if } \Delta(D) \leq \tau \end{cases}$$

GRAM-OOD*

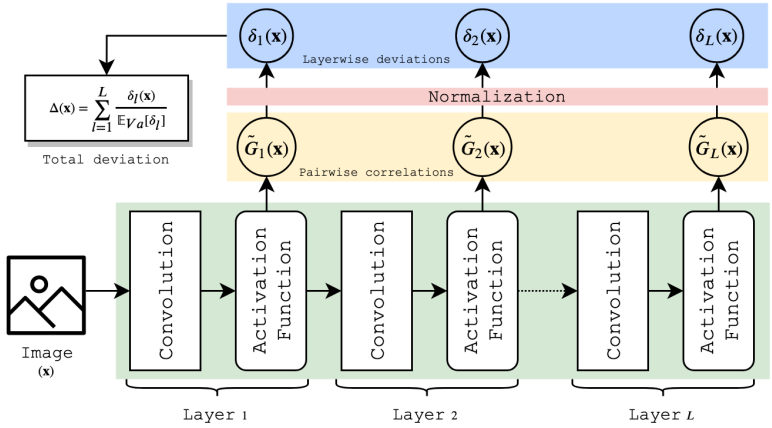
- Normalization of Gram Matrix values

$$\tilde{G}_l^p = \frac{\hat{G}_l^p - \min(\hat{G}_l^p)}{\max(\hat{G}_l^p) - \min(\hat{G}_l^p)}. \tag{3}$$

- Ensures that the class-conditional bounds values are computed from the same interval regardless the layer
- It is possible to consider only activation layers
- It does not require higher-order Gram Matrix for skin cancer detection

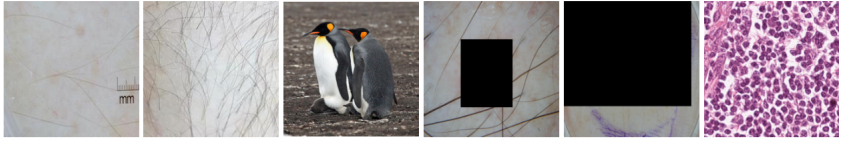
GRAM-OOD*

Overview:



EXPERIMENTS

- In-distributions: ISIC 2019 dataset
- Out-of-distributions: a collection of different datasets



(a) Derm-Skin (b) Clin-Skin (c) Imagenet (d) B-box (e) B-box-70 (f) NCT

- Deep models: DenseNet-121, MobileNet-v2, ResNet-50, and VGGNet-16

EXPERIMENTS

ISIC $\times$ **all**: DenseNet-121 and MobileNet-V2

Model	OOD	TNR @ TPR 95%		
		Mahalanobis (Unbiased)	OOD-Gram	OOD-Gram*
DenseNet-121	Derm-Skin	45.7	**78.0**	76.1
	Clin-Skin	68.6	82.8	**83.1**
	ImageNet	92.0	80.7	**88.4**
	B-box	92.0	88.0	**88.1**
	B-box-70	**100.**	99.9	**100.**
	NCT	91.6	98.9	**99.9**
MobileNet-v2	Derm-Skin	32.4	66.7	**72.8**
	Clin-Skin	79.8	77.9	**83.8**
	ImageNet	85.8	84.3	**92.4**
	B-box	88.4	86.9	**98.7**
	B-box-70	98.4	**100.**	**100.**
	NCT	84.7	99.3	**100.**

EXPERIMENTS

ISIC $\times$ **all**: ResNet-50 and VGGNet-16

Model	OOD	TNR @ TPR 95%		
		Mahalanobis (Unbiased)	OOD-Gram	OOD-Gram*
ResNet-50	Derm-Skin	36.9	**74.8**	73.2
	Clin-Skin	65.9	84.7	**86.3**
	ImageNet	**95.7**	86.6	85.8
	B-box	97.6	88.4	**99.3**
	B-box-70	**100.**	**100.**	**100.**
	NCT	96.9	99.9	**100.**
VGGNet-16	Derm-Skin	31.7	**79.8**	77.5
	Clin-Skin	66.3	80.7	**80.6**
	ImageNet	72.8	77.6	**81.7**
	B-box	85.9	86.5	**94.6**
	B-box-70	93.1	**100**	**100**
	NCT	85.2	99.7	**100.**

EXPERIMENTS

ISIC 2019 Unknown label detection:

Model	AUC	Average Precision
	Mahalanobis / Gram-OOD / Gram-OOD*	
DenseNet-121	52.3 / 67.3 / **69.3**	20.1 / 28.9 / **31.1**
MobileNet-v2	52.9 / 68.7 / **69.5**	20.2 / 31.4 / **32.6**
ResNet-50	56.1 / **70.4** / 70.2	21.6 / 33.2 / **33.7**
VGGNet-16	54.1 / 66.9 / **69.5**	20.9 / 30.2 / **32.6**

CONCLUSION

- Gram-OOD based methods work better than Mahalanobis for the realistic experiment
- Gram-OOD* performs better than the original approach for most of OOD datasets
  - The normalization plays a key role in combining deviations across layers
  - A good normalizing scheme can yield significant improvements in detection rates and should be explored
- Future research: train models that can implicitly detect out-of-distribution samples by taking into account the information contained in the various orders of gram matrices

ACKNOWLEDGMENTS

We thanks the financial support of:

- Coordination for the Improvement of Higher Education Personnel (CAPES)
- National Council for Scientific and Technological Development (CNPq)
- Foundation for Supporting Research and Innovation in Espírito Santo (FAPES)
- Canadian Institute for Advanced Research (CIFAR)

Thank you for your time!

https://github.com/paaatcha/gram-ood