Appendix: Detecting Out-of-Distribution Examples with Gram Matrices

Chandramouli S. Sastry [1] **Sageev Oore** [1]

A. Schematic Diagram

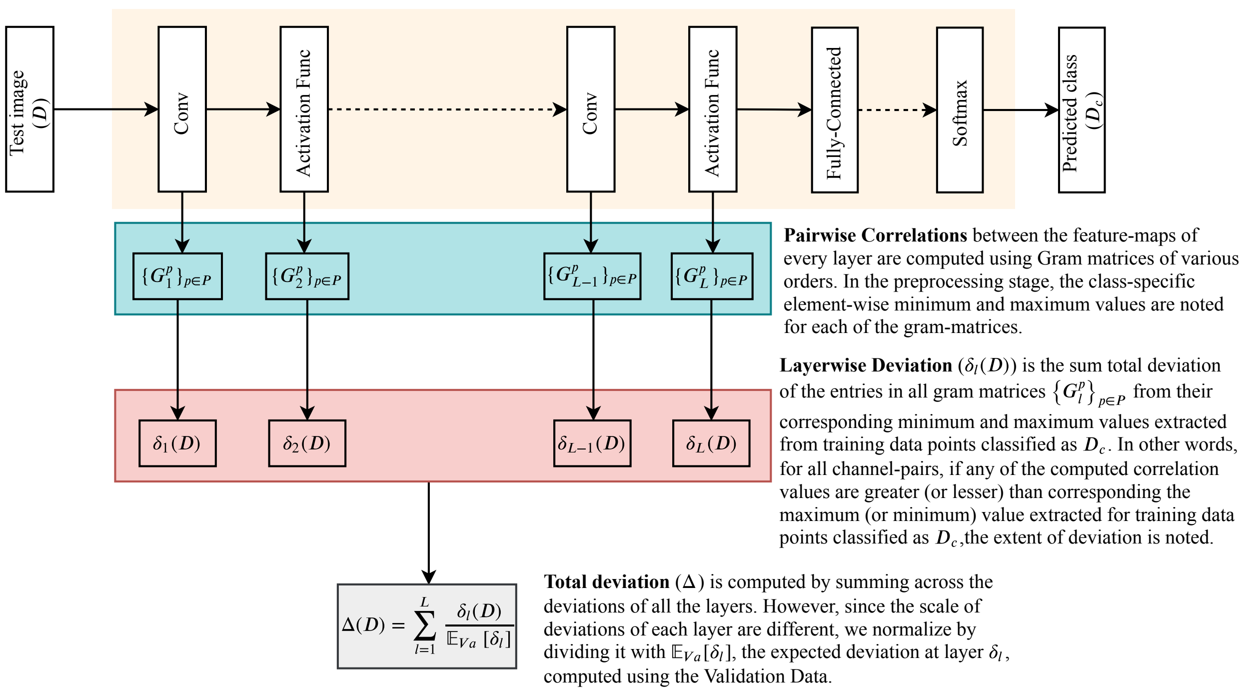

Figure 1: The Schematic Diagram demonstrating the proposed algorithm

B. Description of OOD Datasets

The following includes the description of the out-of-distribution datasets:

1. **TinyImagenet**, a subset of ImageNet (Russakovsky et al., 2015) images, contains 10,000 test images from 200 different classes. Each image is downsampled to size 32 x 32 and all 10,000 images are used, as given in the opensourced version by (Liang et al., 2018).

2. **LSUN**, the Large-scale Scene UNderstanding dataset (Yu et al., 2015) has 10,000 test images from 10 different scenes. Each image is downsampled to size 32 x 32 and all 10,000 images are used, as given in the opensourced version by (Liang et al., 2018).

3. **iSUN**, a subset of SUN images (Xiao et al., 2010), consists of 8925 images. Each image is downsampled to size 32 x 32 and is used; the downsampled version of the dataset has been opensourced by (Liang et al., 2018).

[1]Dalhousie University/ Vector Institute. Correspondence to: Chandramouli Shama Sastry <chandramouli.sastry@gmail.com>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

4. **SVHN**, the Street View House Numbers dataset (Netzer et al., 2011), involves recognizing digits 0-9 in natural scene images. The test partition consisting of 26,032 images is used.

C. Detailed Ablation Results

The results in the main paper correspond to the performance obtained when considering:

1. **Feature Set**: all gram matrix entries
2. **Metric**: layerwise deviations computed with respect to the mins and maxs.
3. **Aggregation Scheme**: the total deviation is then computed using Eq 5.

In this section, detailed ablation results are reported by considering other options. Specifically:

1. **Alternate Feature Set**: In addition to considering all gram matrix entries, we consider a proper partition of the gram matrix: strictly diagonal elements, and strictly off-diagonal elements. The diagonal elements correspond to the unary features, while the off-diagonal elements correspond to pairwise features. This can be done by appropriately changing the definition of variable *stat* in Line 7 of Algorithm 1. In these experiments, we consider row-wise sums wherever the size of stat is $O(n^2)$; in other words, we consider row-wise sums when considering off-diagonal elements and all gram matrix entries.

2. **Alternate Metric**: An alternative formulation for computing feature-wise deviations can be to compute the deviation from the means using the one-dimensional Mahalanobis distance. In the preprocessing stage, this would be done by storing the *Means* and *Variances* of *stat* (feature-wise) instead of their *Mins* and *Maxs*. Under this new alternative, the function δ defined in Eq 3 would be redefined as:

$$\delta(\text{mean,variance,value}) = \frac{(\text{value} - \text{mean})^2}{\text{variance}} \tag{C.1}$$

   Accordingly, the layerwise deviation δ_l can be defined as:

$$\delta_l(D) = \sum_{p=1}^{P} \sum_{i=1}^{|\overline{G_l^p(D)}|} \delta\left(\text{Means}[D_c][l][p][i], \text{Variances}[D_c][l][p][i], \overline{G_l^p(D)}[i]\right) \tag{C.2}$$

   where $\overline{G_l^p(D)}$ would correspond to the statistic chosen in the previous step: diagonal entries only, row-wise sums of off-diagonal entries only or row-wise sums of complete gram matrix.

   We thus consider 2 options for computing the deviations: the Min/Max method presented in the main paper and the Mean/Variance method (Gaussian) described above. While the Mean/Var assumes each entry of the gram matrix to be normally distributed, the Min/Max assumes that each entry is uniformly distributed between the corresponding extrema and has exponentially decreasing density beyond the extrema:

$$p(\text{value}|\text{min, max}) = \begin{cases} k & \text{if min } \leq \text{value} \leq \text{max} \\ k\exp\left(\frac{\text{value}-\text{min}}{|\text{min}|}\right) & \text{if value} < \text{min} \\ k\exp\left(\frac{\text{max}-\text{value}}{|\text{max}|}\right) & \text{if value} > \text{max} \end{cases} \tag{C.3}$$

   where, $k = \frac{1}{max-min+\frac{1}{|min|}+\frac{1}{|max|}}$ makes the above a valid probability density function. Note that both of the proposed density models assume that the entries of the gram matrix are independent of each other. The corresponding δ_l can be obtained as the sum of the log probability density estimates.

3. **Alternate Aggregation Scheme**: In order to compute the total deviation Δ from the layerwise deviations δ_l, we can compute it by following Eq 5 or taking a simple sum as shown:

$$\Delta(D) = \sum_{l=1}^{L} \delta_l \tag{C.4}$$

   We refer to Eq 5 as the normalized estimate and Eq C.4 as the unnormalized estimate.

In all, the Table 1 reports detection rates in 12 settings: 3 choices for *stat* (only the diagonal entries of gram matrix G, only the off-diagonal entries of G, or all of G) $\times$ 2 metrics for computing deviation (Min/Max or Mean/Variance) $\times$ 2 choices for computing total deviation (Normalized sum or Unnormalized sum). All layers and all orders of gram matrix are considered in Table 4.

In-dist (model)	OOD	TNR at TPR95											
		Diagonal Elements				Off Diagonal Elements (Row-wise Sums)				Complete Gram Matrix (Row-wise Sums)			
		Min/ Max	Mean/ Var	Min/ Max (U)	Mean/ Var (U)	Min/ Max	Mean/ Var	Min/ Max (U)	Mean/ Var (U)	Min/ Max	Mean/ Var	Min/ Max (U)	Mean/ Var (U)
CIFAR-10 (ResNet)	iSUN	99.1	96.3	95.7	73.8	99.3	97.1	97.0	90.2	99.3	97.1	97.1	89.3
	LSUN	99.5	98.3	97.0	77.7	99.6	98.8	98.1	93.6	99.6	98.8	98.0	94.1
	TinyImgNet	98.6	93.8	96.2	68.2	98.8	97.2	95.7	88.5	98.7	97.2	95.7	87.5
	SVHN	97.8	70.7	94.8	19.9	97.6	81.1	94.8	40.7	97.6	80.2	94.9	38.2
	CIFAR-100	33.3	29.4	42.5	32.9	32.9	27.2	41.8	29.3	32.9	27.4	42.2	29.2
CIFAR-100 (ResNet)	iSUN	93.8	71.8	50.0	33.8	95.4	85.9	67.2	55.0	94.8	85.3	65.4	52.8
	LSUN	95.6	70.8	45.6	32.9	97.2	87.5	64.2	52.0	96.6	86.8	62.4	49.7
	TinyImgNet	94.1	68.0	51.4	34.6	95.3	84.2	68.1	52.8	94.8	83.5	66.6	50.8
	SVHN	83.1	29.8	53.2	26.0	79.1	34.4	51.8	29.6	80.8	33.9	55.6	29.2
	CIFAR-10	12.9	18.1	19.2	18.2	11.4	17.5	17.5	17.9	12.2	17.6	18.1	18.0
CIFAR-10 (DenseNet)	iSUN	98.9	97.1	98.8	96.3	99.1	97.8	99.1	97.5	99.0	97.8	99.0	97.5
	LSUN	99.4	98.8	99.4	98.4	99.5	99.2	99.5	99.0	99.5	99.1	99.4	98.9
	TinyImgNet	98.7	97.5	98.6	97.0	98.8	98.0	98.8	97.7	98.8	97.9	98.7	97.7
	SVHN	96.6	87.8	96.9	88.2	95.9	84.7	96.4	86.9	96.1	84.0	96.5	87.1
	CIFAR-100	26.4	28.9	29.0	28.2	27.0	25.2	30.1	24.8	26.7	25.5	30.1	25.1
CIFAR-100 (DenseNet)	iSUN	96.0	84.2	95.5	91.9	96.1	88.8	95.5	95.7	95.9	88.5	95.3	95.8
	LSUN	97.4	87.5	96.9	95.5	97.5	91.4	97.0	97.8	97.2	91.2	96.8	97.8
	TinyImgNet	95.8	81.4	95.4	90.2	95.9	86.9	95.3	94.2	95.7	86.4	95.2	94.3
	SVHN	89.4	59.7	88.8	64.5	87.3	63.2	86.4	67.4	89.3	62.9	87.9	67.3
	CIFAR-10	10.5	16.4	11.1	13.7	10.6	15.6	10.2	13.9	10.6	15.6	10.2	13.8
SVHN (ResNet)	iSUN	99.3	99.8	98.1	96.9	99.5	99.9	97.8	98.5	99.4	99.9	97.9	98.6
	LSUN	99.6	99.9	98.5	97.4	99.6	99.9	98.3	99.0	99.6	99.9	98.4	99.0
	TinyImgNet	99.0	99.6	97.8	96.5	99.3	99.6	97.8	98.1	99.3	99.6	98.4	99.0
	CIFAR-10	82.6	93.7	86.1	86.8	85.9	94.9	86.5	90.5	85.8	94.9	86.6	90.5
SVHN (DenseNet)	iSUN	99.3	99.6	99.4	99.1	99.3	99.4	99.4	97.9	99.4	99.4	99.4	98.0
	LSUN	99.5	99.7	99.7	99.4	99.5	99.4	99.6	98.2	99.5	99.4	99.6	98.3
	TinyImgNet	99.2	99.5	99.3	99.2	99.1	99.2	99.2	98.0	99.1	99.2	99.2	98.1
	CIFAR-10	76.6	93.5	76.9	91.8	81.2	94.3	85.6	90.1	80.4	94.2	84.7	90.1
Summary	MEAN	95.4	86.6	87.9	77.3	95.6	90.1	90.4	83.7	95.7	89.9	90.4	83.3
	STD-DEV	6.2	17.2	18.1	27.0	6.0	14.7	13.5	21.3	5.7	14.9	13.5	22.0

In-dist (model)	OOD	AUROC											
		Diagonal Elements				Off Diagonal Elements (Row-wise Sums)				Complete Gram Matrix (Row-wise Sums)			
		Min/ Max	Mean/ Var	Min/ Max (U)	Mean/ Var (U)	Min/ Max	Mean/ Var	Min/ Max (U)	Mean/ Var (U)	Min/ Max	Mean/ Var	Min/ Max (U)	Mean/ Var (U)
CIFAR-10 (ResNet)	iSUN	99.7	99.2	98.7	96.2	99.7	99.4	99.0	98.1	99.8	99.4	99.0	98.0
	LSUN	99.8	99.5	98.7	96.7	99.8	99.7	99.1	98.5	99.9	99.7	99.1	98.4
	TinyImgNet	99.6	99.1	98.3	95.6	99.6	99.4	98.8	97.8	99.7	99.4	98.8	97.7
	SVHN	99.5	95.2	98.9	88.4	99.4	96.2	98.8	92.0	99.5	96.2	98.9	91.8
	CIFAR-100	79.7	78.4	84.9	83.3	78.8	75.8	84.1	79.0	79.0	76.1	84.2	79.2
CIFAR-100 (ResNet)	iSUN	98.7	95.5	92.1	87.8	98.9	97.4	94.5	92.7	98.8	97.3	94.2	92.3
	LSUN	99.1	95.9	91.8	87.8	99.3	97.8	94.6	92.9	99.2	97.7	94.3	92.4
	TinyImgNet	98.8	95.1	92.5	87.1	99.0	97.2	94.8	92.5	98.9	97.1	94.6	92.1
	SVHN	96.5	84.7	92.3	80.8	95.7	86.8	91.6	83.5	96.0	86.7	92.2	83.1
	CIFAR-10	69.3	71.2	76.6	74.6	67.3	70.0	75.3	71.9	67.9	70.1	75.5	72.0
CIFAR-10 (DenseNet)	iSUN	99.8	99.5	99.8	99.3	99.8	99.6	99.8	99.6	99.8	99.6	99.8	99.5
	LSUN	99.9	99.8	99.9	99.7	99.9	99.8	99.9	99.8	99.9	99.8	99.9	99.8
	TinyImgNet	99.7	99.5	99.7	99.3	99.7	99.6	99.7	99.5	99.7	99.6	99.7	99.5
	SVHN	99.2	97.6	99.3	97.6	99.1	96.9	99.2	97.4	99.1	96.8	99.2	97.4
	CIFAR-100	68.5	75.1	68.9	74.1	72.1	72.7	73.3	72.4	72.0	72.9	73.4	72.5
CIFAR-100 (DenseNet)	iSUN	99.1	97.2	98.9	98.2	99.0	97.9	98.9	99.0	99.0	97.8	98.9	99.0
	LSUN	99.4	97.8	99.3	99.0	99.4	98.3	99.3	99.4	99.3	98.3	99.3	99.4
	TinyImgNet	99.0	96.6	98.9	97.8	99.0	97.5	98.9	98.7	99.0	97.4	98.9	98.7
	SVHN	97.4	92.5	97.3	92.6	97.0	92.7	96.9	93.4	97.3	92.7	97.1	93.4
	CIFAR-10	64.4	70.1	65.0	66.7	63.7	68.3	64.2	66.2	64.2	68.7	64.6	66.2
SVHN (ResNet)	iSUN	99.8	99.9	98.8	98.3	99.8	99.9	99.0	99.0	99.8	99.9	99.0	99.0
	LSUN	99.9	99.9	98.9	98.4	99.8	99.9	99.1	99.1	99.8	99.9	99.1	99.1
	TinyImgNet	99.7	99.8	98.8	98.3	99.7	99.8	99.0	99.0	99.7	99.8	99.0	99.0
	CIFAR-10	96.8	98.6	97.1	97.3	97.3	98.8	97.3	97.8	97.3	98.8	97.3	97.8
SVHN (DenseNet)	iSUN	99.8	99.9	99.9	99.8	99.8	99.9	99.8	99.5	99.8	99.9	99.8	99.5
	LSUN	99.9	99.9	99.9	99.9	99.8	99.9	99.9	99.6	99.8	99.9	99.9	99.6
	TinyImgNet	99.7	99.9	99.8	99.8	99.7	99.8	99.8	99.6	99.7	99.8	99.8	99.6
	CIFAR-10	94.5	98.5	94.9	98.1	95.6	98.6	96.5	97.6	95.5	98.6	96.4	97.7
Summary	MEAN	99.0	97.5	97.7	95.6	99.0	98.0	98.1	96.9	99.0	98.0	98.1	96.8
	STD-DEV	1.3	3.4	2.7	5.2	1.3	2.9	2.2	3.9	1.2	2.9	2.2	4.0

In-dist (model)	OOD	DTACC											
		Diagonal Elements				Off-Diagonal Elements (Row-wise Sums)				Complete Gram Matrix (Row-wise Sums)			
		Min/ Max	Mean/ Var	Min/ Max (U)	Mean/ Var (U)	Min/ Max	Mean/ Var	Min/ Max (U)	Mean/ Var (U)	Min/ Max	Mean/ Var	Min/ Max (U)	Mean/ Var (U)
CIFAR-10 (ResNet)	iSUN	97.9	95.7	95.5	92.1	98.0	96.2	96.0	93.5	98.1	96.2	96.0	93.4
	LSUN	98.5	97.0	96.0	93.0	98.6	97.6	96.6	94.6	98.6	97.6	96.6	94.5
	TinyImgNet	97.5	95.6	94.7	91.2	97.7	96.3	95.3	93.1	97.8	96.4	95.4	92.9
	SVHN	97.0	90.1	95.0	84.7	96.6	90.9	94.9	87.0	96.7	90.8	95.1	86.7
	CIFAR-100	72.4	71.7	78.2	76.8	71.5	69.1	77.4	72.1	71.7	69.4	77.4	72.2
CIFAR-100 (ResNet)	iSUN	94.5	89.3	85.4	81.1	95.3	92.2	87.9	86.1	95.6	92.0	87.6	85.6
	LSUN	95.4	90.1	85.7	81.4	96.3	93.1	88.4	86.6	96.7	93.0	88.1	86.1
	TinyImgNet	94.6	88.5	85.9	80.1	95.2	91.8	88.4	85.9	95.0	91.6	88.2	85.4
	SVHN	90.2	77.5	85.1	73.6	89.2	80.3	84.3	75.3	89.6	80.3	84.8	74.9
	CIFAR-10	64.6	66.0	71.1	69.1	63.0	64.8	69.8	66.5	63.4	65.0	70.1	66.6
CIFAR-10 (DenseNet)	iSUN	97.8	96.5	97.8	95.9	97.9	96.9	97.8	96.8	97.9	96.8	97.8	96.7
	LSUN	98.7	98.0	98.6	97.4	98.6	98.2	98.5	98.1	98.6	98.1	98.5	98.1
	TinyImgNet	97.9	96.8	97.8	96.3	97.8	97.1	97.7	97.0	97.9	97.0	97.7	97.0
	SVHN	96.3	92.3	96.5	92.3	95.8	91.1	96.0	92.0	95.9	90.9	96.1	92.0
	CIFAR-100	66.5	68.8	66.6	67.8	67.6	67.2	68.9	66.6	67.3	67.3	68.6	66.6
CIFAR-100 (DenseNet)	iSUN	95.7	91.4	95.4	93.6	95.7	92.6	95.3	95.5	95.6	92.6	95.3	95.5
	LSUN	96.4	92.7	96.1	95.3	96.5	93.7	96.2	96.7	96.4	93.7	96.2	96.7
	TinyImgNet	95.5	90.4	95.2	92.9	95.5	91.8	95.2	94.7	95.5	91.7	95.2	94.7
	SVHN	92.4	85.7	92.0	86.0	91.7	86.0	91.4	87.1	92.4	86.2	91.9	87.1
	CIFAR-10	60.6	64.9	61.3	62.2	59.7	63.4	60.5	61.6	60.4	63.8	61.0	61.5
SVHN (ResNet)	iSUN	98.0	98.4	96.6	96.3	98.1	98.5	96.4	96.8	98.1	98.5	96.5	96.8
	LSUN	98.5	98.9	96.8	96.5	98.5	98.9	96.7	97.0	98.5	98.9	96.7	97.0
	TinyImgNet	97.6	98.1	96.4	96.2	97.9	98.2	96.4	96.6	97.9	98.2	96.5	96.6
	CIFAR-10	91.1	94.9	92.5	94.0	92.0	95.2	92.8	94.3	92.0	95.2	92.8	94.2
SVHN (DenseNet)	iSUN	98.3	98.8	98.4	98.3	98.4	98.5	98.4	97.3	98.3	98.6	98.4	97.3
	LSUN	98.6	98.9	98.7	98.5	98.6	98.6	98.7	97.6	98.6	98.7	98.7	97.7
	TinyImgNet	97.9	98.5	98.1	98.2	98.0	98.3	98.1	97.3	97.9	98.3	98.1	97.3
	CIFAR-10	88.1	94.3	88.6	93.5	89.2	94.7	90.7	92.6	89.1	94.7	90.6	92.6
Summary	MEAN	96.0	93.7	94.1	91.6	96.1	94.4	94.5	92.9	96.2	94.4	94.5	92.8
	STD-DEV	2.9	5.2	4.5	6.8	2.8	4.5	3.9	5.6	2.8	4.5	3.9	5.7

Table 1: Detailed Ablation Results demonstrating the detection rates under 12 different settings. The MEAN and STD-DEV are computed by using all elements in the table excepting the CIFAR-10 vs CIFAR-100 and CIFAR-100 vs CIFAR-10 entries.

By analysing the ablation results, we attempt to answer the following questions:

1. **Are pairwise features more useful than unary features?** We observe that the Min/Max metric can work equally well with both unary and pairwise features; in some cases, the unary features are marginally better (Ex: ResNet/CIFAR-10 vs SVHN) and in some cases, the pairwise features are marginally better (Ex: ResNet/CIFAR-100 vs iSUN/LSUN/TinyImgNet). Interestingly, the behavior of the Mean/Var metric is different: the performance with pairwise features are significantly higher than with unary features in 19 out of 28 tested cases. For example, the TNR at TPR95 for ResNet/CIFAR-100 vs TinyImgNet is 68.0 with unary features and 84.2 with pairwise features.

   We notice that using the unary features (diagonal entries) sometimes did well when pairwise features (off-diagonal entries) did not do well, and vice versa, so using both gives the kind of effect that we *want* in an ensemble: models that cover and work well over different parts of the space. Therefore, an overall message of our experiments is that it is worthwhile to consider all elements of the gram matrix.

2. **Is it neccessary to use Min/Max metric?** Except in 6 cases (ResNet: CIFAR-10 vs CIFAR-100, ResNet: CIFAR-100 vs CIFAR-10, DenseNet: CIFAR-10 vs CIFAR-100, DenseNet: CIFAR-100 vs CIFAR-10, ResNet: SVHN vs CIFAR-10 and DenseNet: SVHN vs CIFAR-10), the min/max metric consistently performs better than the mean/var metric. Additionally, it is not clear if the Mean/Var estimate performs better with normalized sums or unnormalized sums: for example, observe that Mean/Var estimate performs very poorly with the unnormalized estimate for ResNet/CIFAR-100, while the performance of Mean/Var for DenseNet/CIFAR-100 is competitive with the performance of Min/Max only when an unnormalized estimate is computed.

   One can observe that computing the one-dimensional Mahalanobis distance for each component of the statistic derived from the Gram Matrix and later computing the total sum is equivalent to representing each input image by a big vector (say, Z) derived from the Gram Matrices computed across various layers, constructing class-conditional distributions of Z (assuming that each component of Z is normally distributed and independent of the other components) and subsequently computing the probability of an unseen Z. In early research, we noticed the following problems with the Mean/Var estimate:

   (a) The individual components of gram matrices do not follow normal distribution strictly and Mean/Var assigns lower probabilities to the in-distribution images as well.

   (b) The total deviation Δ – computed by simply summing across the layerwise deviations, δ_l – was not able to accurately summarize the information contained in the different δ_ls. Specifically, information about the layer where the input example had a higher deviation was lost when a simple sum was taken.

The proposed Min/Max idea solves problem (a) by employing a weaker metric: deviation from extrema instead of the mean. It can also be said that the Min/Max metric considers a uniform probability density between the extrema. Problem (b), which exists even for this newer metric, is solved by the normalization scheme described in the main paper for computing the sum total deviation.

Higher Order Gram Matrices The Min/Max metric is a weak approximation to the true probability density. On conducting a thorough analysis of how the OOD examples were able to fool the metric, it appeared that the intermediate features had several tiny activations that could yield innocuous gram matrix entries. For example, observe in Fig. 2 of the main paper that the Min/Max metric already gets a detection rate close to that of Mahalanobis by using just Order-1 Gram Matrices. Higher-order gram matrices as described in the main paper provide a natural way to mitigate these effects. More importantly, they help in obtaining descriptive summaries of the high-dimensional feature representations through the higher-order non-central moments – of channels and inter-channel hadamard products – contained in them

Notable observations from Figures 2 through 7 (all layers are considered but only one order of gram matrix is considered at a time):

- **Ensemble effect**: In 24/28 cases, higher order gram matrices improve detection rates. Higher order gram matrices help both the Min/Max and the Mean/Var metrics. In most cases, the even powers are more helpful than the odd powers; in some cases, the odd powers are more helpful (Ex: DenseNet/CIFAR-100 vs CIFAR-10). Despite these variations, it is possible to get an ensemble effect by considering all possible powers as demonstrated in the main paper.
- In ResNet:CIFAR-10 vs CIFAR-100 and DenseNet:CIFAR-10 vs CIFAR-100, the higher order gram matrices yield lower detection rates. We find these exceptions interesting, and would like to understand them better in future.

Summary The unambiguous message from this ablation study is that the Gram matrix contains useful information which can be used for detecting OOD examples. While the standard Mean/Variance metric does not always work well, the proposed Min/Max metric yields consistent performance competitive with state-of-the-art methods. The use of higher-order Gram matrices further boosts the overall performance. Although the Min/Max method can work very well for "far-from-distribution" examples, it does not work well when a fine grained estimate is needed (for example, CIFAR-10 vs CIFAR-100). We hope the strong empirical proof that Gram matrices contain useful information can motivate the development of OOD detectors with powerful density estimators.

C.1. Importance of higher-order Gram Matrices

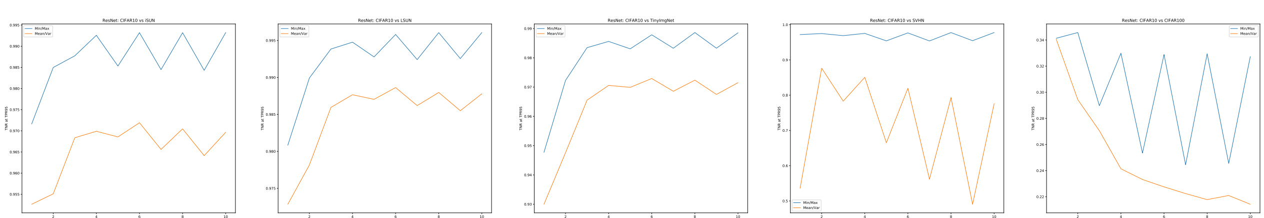

Figure 2: ResNet/CIFAR-10: The TNR at TPR95 trends for Min/Max and Mean/Var as the order of Gram Matrix is varied.

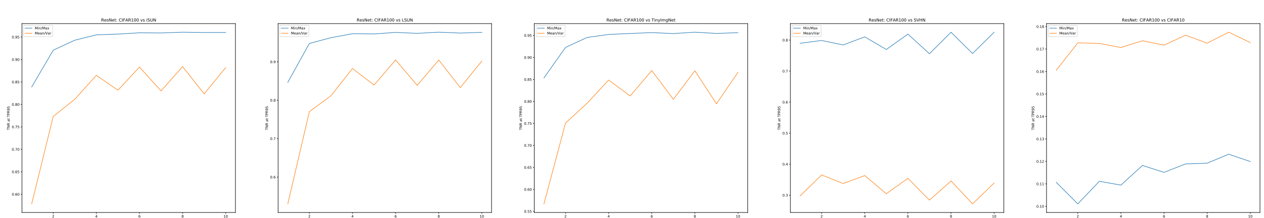

Figure 3: ResNet/CIFAR-100: The TNR at TPR95 trends for Min/Max and Mean/Var as the order of Gram Matrix is varied.

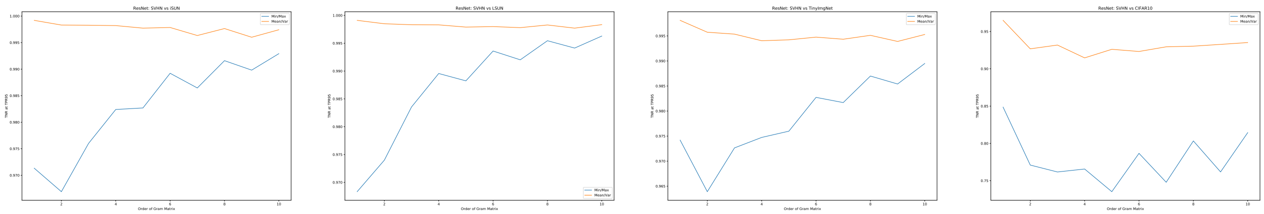

Figure 4: ResNet/SVHN: The TNR at TPR95 trends for Min/Max and Mean/Var as the order of Gram Matrix is varied.

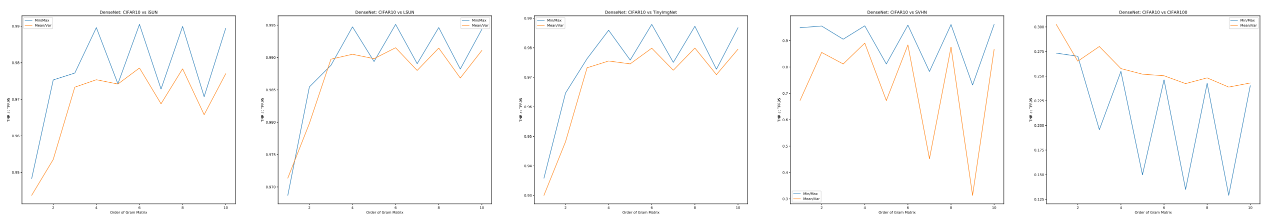

Figure 5: DenseNet/CIFAR-10: The TNR at TPR95 trends for Min/Max and Mean/Var as the order of Gram Matrix is varied.

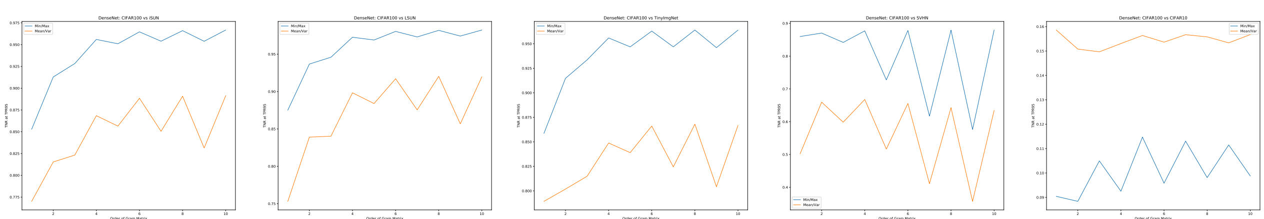

Figure 6: DenseNet/CIFAR-100: The TNR at TPR95 trends for Min/Max and Mean/Var as the order of Gram Matrix is varied.

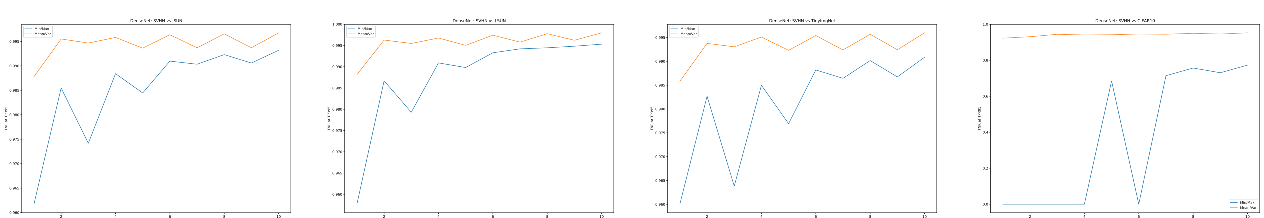

Figure 7: DenseNet/SVHN: The TNR at TPR95 trends for Min/Max and Mean/Var as the order of Gram Matrix is varied.

C.2. Significance of Depth

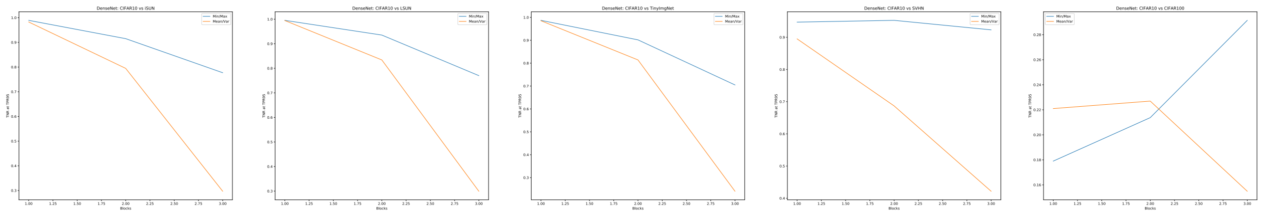

Figure 8: DenseNet/CIFAR-10: The TNR at TPR95 trends for Min/Max and Mean/Var as we go deeper in the network.

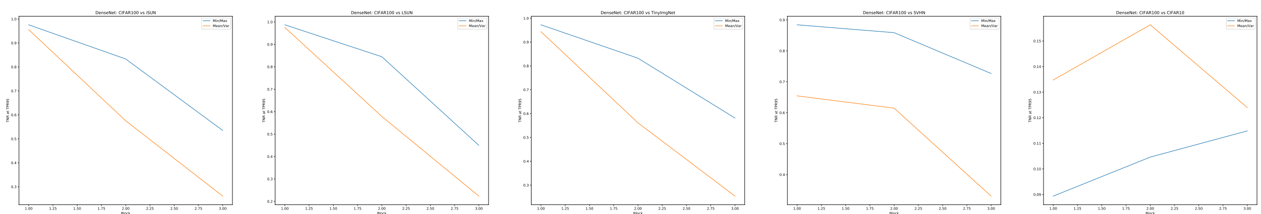

Figure 9: DenseNet/CIFAR-100: The TNR at TPR95 trends for Min/Max and Mean/Var as we go deeper in the network.

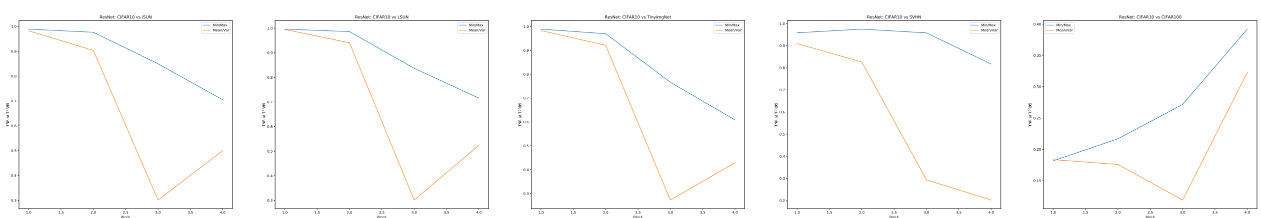

Figure 10: ResNet/CIFAR-10: The TNR at TPR95 trends for Min/Max and Mean/Var as we go deeper in the network.

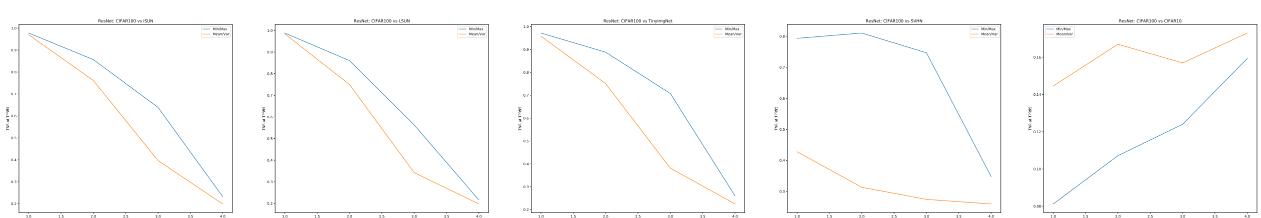

Figure 11: ResNet/CIFAR-100: The TNR at TPR95 trends for Min/Max and Mean/Var as we go deeper in the network.

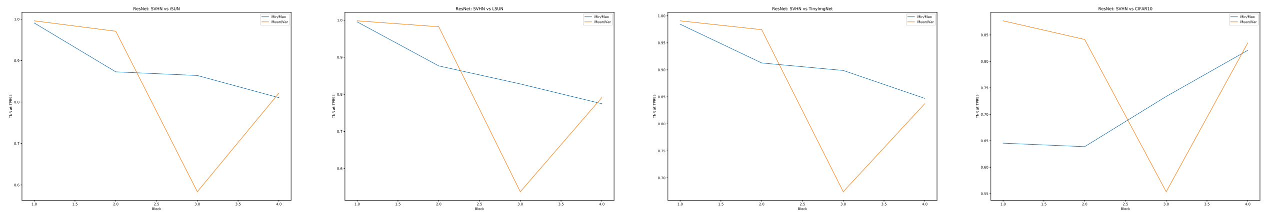

Figure 12: DenseNet/SVHN: The TNR at TPR95 trends for Min/Max and Mean/Var as we go deeper in the network.

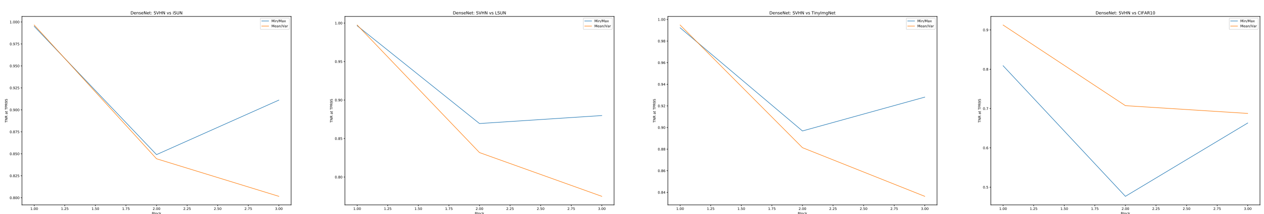

Figure 13: DenseNet/SVHN: The TNR at TPR95 trends for Min/Max and Mean/Var as we go deeper in the network.

D. Combining OE + Ours

In-dist (WRN 40-2)	OOD	MSP			Ours			Ours + MSP		
		TNR at TPR 95%	AUROC	DTACC	TNR at TPR 95%	AUROC	DTACC	TNR at TPR 95%	AUROC	DTACC
CIFAR-10	iSUN	98.3	99.3	96.9	98.9	99.8	97.8	**99.8**	99.9	99.0
	LSUN (R)	98.5	99.4	97.0	99.4	99.9	98.4	**99.8**	99.9	99.1
	LSUN (C)	98.0	99.4	96.9	89.5	97.8	92.5	**98.6**	99.6	97.3
	TinyImgNet (R)	93.9	98.5	94.6	98.5	99.7	97.6	**99.5**	99.9	98.5
	TinyImgNet (C)	95.2	98.7	95.2	95.9	99.1	95.7	**99.1**	99.8	97.8
	SVHN	98.0	99.5	96.9	97.6	99.4	96.8	**99.3**	99.8	98.2
	CIFAR-100	**73.9**	94.8	87.9	38.9	80.1	73.3	72.9	93.9	87.0
CIFAR-100	iSUN	50.9	89.8	82.3	**96.3**	99.1	95.9	95.6	98.9	96.0
	LSUN (R)	58.3	92.0	84.7	**98.4**	99.6	97.3	97.4	99.3	97.4
	LSUN (C)	69.5	94.0	86.6	69.7	92.6	85.3	**83.1**	96.3	89.7
	TinyImgNet (R)	36.1	85.1	77.5	**96.3**	99.1	95.9	92.8	98.2	94.6
	TinyImgNet (C)	41.6	86.3	78.6	**90.1**	97.7	92.8	87.1	96.9	91.1
	SVHN	56.2	92.5	85.6	84.8	96.5	90.8	**85.6**	96.8	90.4
	CIFAR-10	**17.4**	78.4	71.7	7.5	59.3	57.3	16.5	77.7	71.6

Table 2: Table shows results when our method is combined with OE. The experiment was conducted with WideResNet trained with outlier-exposure, open-sourced by (Hendrycks et al., 2019). MSP uses Maximum Softmax Probability; "Ours" refers to the metric Δ (Eq. 5); "Ours+MSP" is obtained by using $\Delta'(x) = \frac{\Delta(x)}{\max_{x \in \mathrm{Va}} \Delta(x)} - \text{MSP}$.

E. Few more OOD results

E.1. Comparing with OE

In-distribution	OOD	OE (Base)	OE	Ours (Base)	Ours
CIFAR-10	Gaussian	85.6	**99.3**	43.5	**100.**
	Rademacher	52.4	**99.5**	48.3	**100.**
	Blob	83.8	**99.4**	52.9	**99.8**
	Texture	57.2	**87.8**	37.0	85.3
	SVHN	71.2	95.2	45.4	**96.1**
	LSUN	61.3	87.9	58.2	**99.5**
CIFAR-100	Gaussian	45.7	87.9	18.2	**100.**
	Rademacher	61.0	82.9	15.6	**100.**
	Blob	62.0	87.9	38.4	**98.6**
	Texture	28.5	45.6	19.9	**68.5**
	SVHN	30.7	57.1	23.5	**85.4**
	LSUN	26.0	42.5	18.2	**97.2**
SVHN	Gaussian	94.6	**100.**	87.65	**100.**
	Bernoulli	95.6	**100.**	92.25	**100.**
	Blob	96.3	**100.**	93.35	**100.**
	Texture	92.8	**99.8**	72.6	94.9
	Cifar-10	94.0	**99.9**	73.8	83.0
	LSUN	93.6	**99.9**	75.7	**99.5**

Table 3: Comparison of Mean TNR@TPR95 values.

Following (Hendrycks et al., 2019), we created the gaussian, rademacher, blob and bernoulli synthetic datasets. Their descriptions are as follows: *Gaussian* anomalies have each dimension i.i.d. sampled from an isotropic Gaussian distribution. *Rademacher* anomalies are images where each dimension is -1 or 1 with equal probability, so each dimension is sampled from a symmetric Rademacher distribution. *Bernoulli* images have each pixel sampled from a Bernoulli distribution if the input range is [0, 1]. *Blobs* data consist of algorithmically generated amorphous shapes with definite edges. *Textures* is a dataset of describable textural images (Cimpoi et al., 2014).

E.2. Comparing with DPN, VD and Semantic.

OOD	Method	TNR @ TPR95	AUROC	Detection Accuracy
LSUN	DPN	42.60	90.20	79.50
	VD	92.30	98.30	94.10
	Baseline	49.80	91.00	85.30
	ODIN	82.10	94.10	86.70
	Mahalanobis	98.80	99.70	97.70
	Ours	**99.85**	**99.89**	**98.66**
Tiny ImgNet	DPN	71.60	93.00	86.40
	VD	82.90	96.80	91.30
	Baseline	41.00	91.00	85.10
	ODIN	67.90	94.00	86.50
	Mahalanobis	97.10	99.50	96.30
	Ours	**99.48**	**99.72**	**97.82**
SVHN	DPN	79.90	95.90	87.30
	VD	71.30	93.20	86.40
	Baseline	50.50	89.90	85.10
	ODIN	70.30	96.70	91.10
	Mahalanobis	87.80	99.10	95.80
	Ours	**98.14**	**99.50**	**96.71**

(a) ResNet/CIFAR-10

OOD	Method	TNR @ TPR95	AUROC	Detection Accuracy
iSUN	Semantic	41.60	85.20	88.40
	VD	80.20	94.20	87.80
	Baseline	16.89	75.80	70.11
	ODIN	45.21	85.48	78.47
	Mahalanobis	89.91	97.91	93.05
	Ours	**95.12**	**98.9**	**95.18**
LSUN	Semantic	20.50	79.00	57.80
	VD	85.50	95.90	90.40
	Baseline	18.80	75.80	69.90
	ODIN	23.20	85.60	78.30
	Mahalanobis	90.89	98.2	93.5
	Ours	**97.14**	**99.28**	**96.19**
Tiny ImgNet	Semantic	37.60	83.10	75.60
	VD	83.70	95.30	89.70
	Baseline	20.40	77.20	70.80
	ODIN	36.1	87.6	80.1
	Mahalanobis	90.92	98.20	93.30
	Ours	**95.12**	**98.97**	**95.13**

(b) ResNet/CIFAR-100

Table 4: We compare our method with DPN, VD and Semantic by reporting results where available.

E.3. Results for Fully-connected Networks

Architecture	OOD	Method	TNR @ TPR95	AUROC	Detection Accuracy
300	KMNIST	Baseline	47.66	73.96	73.91
		Ours	**98.57**	**99.66**	**97.37**
	Fashion-MNIST	Baseline	44.93	66.93	71.07
		Ours	**93.51**	**98.64**	**94.36**
300-150	KMNIST	Baseline	59.79	75.17	79.49
		Ours	**97.8**	**99.4**	**96.55**
	Fashion-MNIST	Baseline	70.73	77.10	83.00
		Ours	**95.2**	**99.00**	**95.17**
300-150-50	KMNIST	Baseline	70.4	79.75	83.38
		Ours	**97.5**	**99.11**	**96.4**
	Fashion-MNIST	Baseline	73.92	76.54	84.67
		Ours	**95.7**	**98.94**	**95.48**

Table 5: The method even works quite well with a fully-connected neural network trained on MNIST. The results are shown for 300-unit single layer MLP, 300-150 two-layer MLP and 300-150-50 MLP.

F. SVHN images

Figure 14: Some images selected from the test partition of SVHN which have unusual feature correlations as determined by our method. Some images we found interesting include what appears to be a porch lamp (Row 2 Col 5) and an 8 inside a 0 (Row 2 Col 3).

References

M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi. Describing textures in the wild. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, pages 3606–3613, 2014. doi: 10.1109/CVPR.2014.461. URL `https://doi.org/10.1109/CVPR.2014.461`.

D. Hendrycks, M. Mazeika, and T. G. Dietterich. Deep anomaly detection with outlier exposure. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019. URL `https://openreview.net/forum?id=HyxCxhRcY7`.

S. Liang, Y. Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018. URL `https://openreview.net/forum?id=H1VGkIxRZ`.

Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011*, 2011. URL `http://ufldl.stanford.edu/housenumbers/nips2011_housenumbers.pdf`.

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115 (3):211–252, Dec 2015. ISSN 1573-1405. doi: 10.1007/s11263-015-0816-y. URL `https://doi.org/10.1007/s11263-015-0816-y`.

J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. SUN database: Large-scale scene recognition from abbey to zoo. In *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*, pages 3485–3492, 2010. doi: 10.1109/CVPR.2010.5539970. URL `https://doi.org/10.1109/CVPR.2010.5539970`.

F. Yu, Y. Zhang, S. Song, A. Seff, and J. Xiao. LSUN: construction of a large-scale image dataset using deep learning with humans in the loop. *CoRR*, abs/1506.03365, 2015. URL `http://arxiv.org/abs/1506.03365`.