IEEE ICIP 2022

Subspace Modeling For Fast Out-of-Distribution and Anomaly Detection

Ibrahima J. Ndiour, Nilesh Ahuja and Omesh Tickoo

Sr. Research Scientists, Intel Labs

intel.

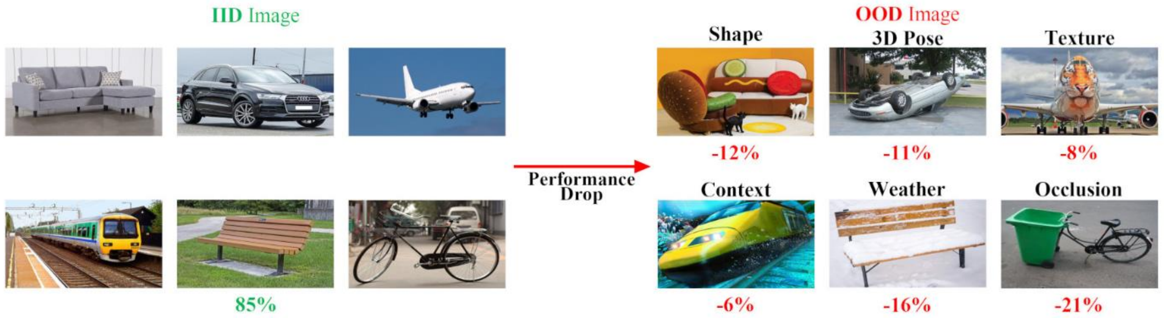
ICIP2022 BORDEAUX

IEEE Signal Processing Society

Introduction to OOD and anomaly detection

Out-of-Distribution detection

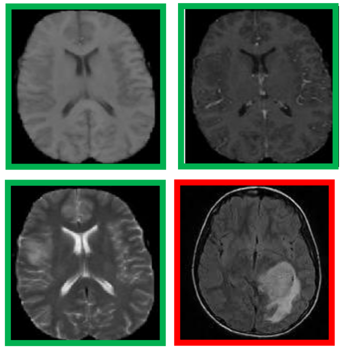
- OOD: new, previously unseen data that does not resemble training data.
- May lead to unpredictable and sometimes catastrophic outputs.
- Important for uncertainty estimation in AI, active learning, explainability and dataset bias.



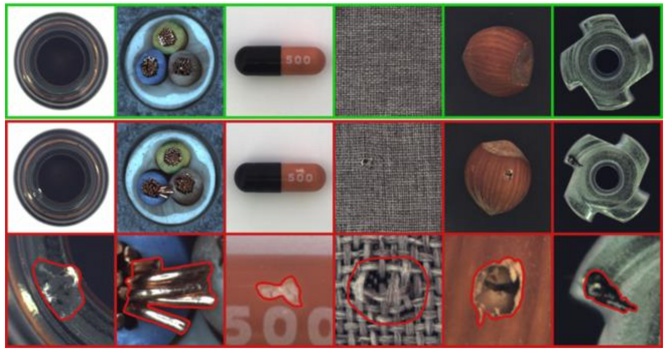
OOD-CV challenge dataset

Anomaly detection

- The goal is to identify rare and abnormal events/samples from data observation.



Brain MRI data samples



MVTec-AD dataset

Proposed Approach

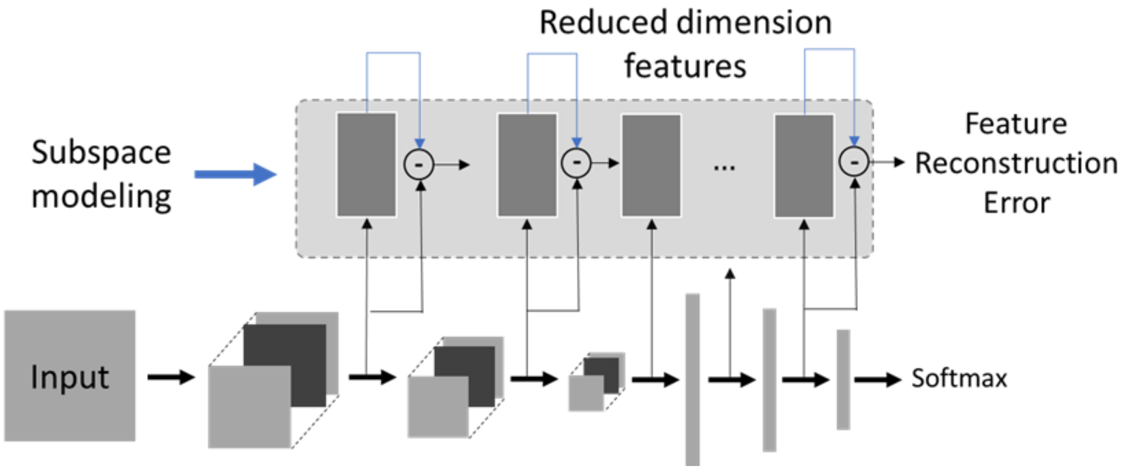
At training:

- Learn a transformation $\mathcal{T}$ that maps the high-dimensional features onto an appropriate subspace, $\mathcal{T} : \mathcal{H} \rightarrow \mathcal{L}$ with $dim(\mathcal{L}) \ll dim(\mathcal{H})$, along with the inverse transformation $\mathcal{T}^{\dagger}$.

During inference:

- Apply the transformation to a test feature x, then inverse-transform it back into the original space and calculate the feature reconstruction error (FRE) as:

$$FRE(\mathbf{x}) = \|f(\mathbf{x}) - (\mathcal{T}^{\dagger} \circ \mathcal{T})(f(\mathbf{x}))\|_2$$

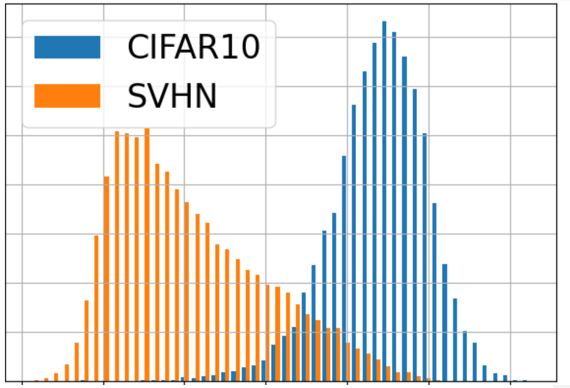


Proposed system

Experimental validation: setup and metrics

OOD and anomaly detection:

- FOR OOD, choose an in-distribution dataset and a given CNN backbone (e.g. CIFAR100 and Wide-Resnet):
  - Analysis is performed on 3 given layers.
  - Use 2 other datasets for OOD data (e.g. SVHN, LSUN).
- For AD, use defective samples from MVTec and MTD.
- Estimate the subspace model parameters on the training split of the in-distribution dataset (both PCA and kPCA with RBF kernel).
- Evaluate quality performance on the test splits.
  - FRE is used as uncertainty score.
- Use AUROC metric to benchmark our OOD classifier's performance



Experimental validation: OOD detection

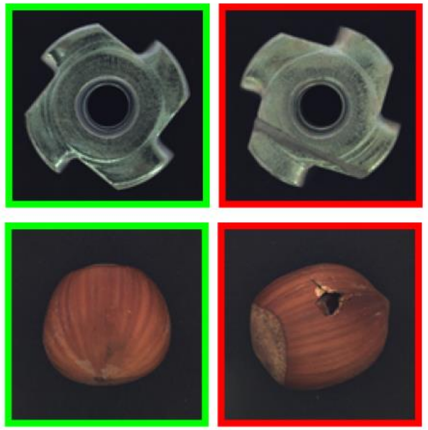
	Mahal	LL	FRE	kFRE	Mahal	LL	FRE	kFRE
CIFAR100	**SVHN (OOD)**				**LSUN (OOD)**			
Layer2	91.5	92.8	**93.1**	**93.4**	**98.5**	**98.8**	**98.3**	98.2
Layer1	91.2	90.0	93.2	**95.8**	**98.7**	**99.1**	98.4	98.3
Layer0	75.0	**84.8**	79.3	75.9	**97.3**	95.1	**97.1**	91.5
Softmax	74.3				84.7			
CIFAR10	**SVHN (OOD)**				**LSUN (OOD)**			
Layer2	94.6	94.5	77.2	**98.5**	98.8	**99.4**	95.3	**99.0**
Layer1	86.4	88.8	48.5	**92.4**	72.5	86.0	65.2	**87.0**
Layer0	95.2	95.0	**96.7**	93.9	**95.1**	**95.5**	**95.3**	**95.1**
Softmax	93.4				94.0			
SVHN	**CIFAR10 (OOD)**				**LSUN (OOD)**			
Layer2	**94.2**	**93.8**	85.2	**93.7**	**94.3**	**93.9**	90.1	93.5
Layer1	90.4	**94.9**	94.2	**94.7**	90.6	**95.2**	94.5	**94.9**
Layer0	92.3	**96.8**	96.0	95.6	92.5	**97.1**	96.0	95.8
Softmax	93.0				92.5			

OOD detection AUROC

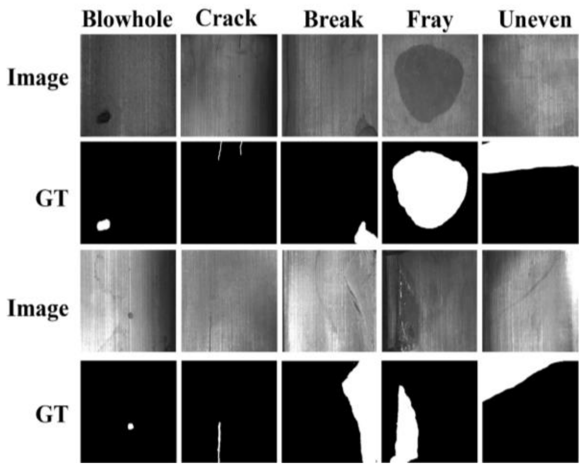
Experimental validation: anomaly detection

Category	GANomaly	DifferNet	SPADE*	PaDim*	PatchCore	FRE (Ours)
Carpet	69.9	92.9	-	99.5	98.7	100
Grid	70.8	84.0	-	94.2	98.2	95.8
Leather	84.2	97.1	-	100	100	100
Tile	79.4	99.4	-	97.4	98.7	97.8
Wood	83.4	99.8	-	99.3	99.2	99.4
Bottle	89.2	99.0	-	99.9	100	100
Cable	75.7	95.9	-	87.8	99.5	99.3
Capsule	73.2	86.9	-	92.7	98.1	99.4
Hazelnut	78.5	99.3	-	96.4	100	99.8
Metal Nut	70.0	96.1	-	98.9	100	96.9
Pill	74.3	88.8	-	93.9	96.6	95.7
Screw	74.6	96.3	-	84.5	98.1	97.5
Toothbrush	65.3	98.6	-	94.2	100	99.4
Transistor	79.2	91.1	-	97.6	100	98.6
Zipper	74.5	95.1	-	88.2	99.4	96.4
Average	**76.2**	**94.9**	**85.5**	**97.9**	**99.1**	**98.4**

Anomaly detection AUROC on MVTec-AD



MVTec-AD samples



Magnetic Tiles samples

GANomaly	I-NN	DifferNet	PatchCore	FRE (Ours)
76.6	80.0	97.7	97.9	99.2

Anomaly detection AUROC on Magnetic Tiles

Model	Inference Frame-rate (fps)	Avg. Training Time (s)
EfficientNet-B5	61.5	7.35

Inference framerate and average training time for a single MVTec category (measured on Intel Xeon 8280 CPU)

Conclusion and Future Work

Out-of-Distribution and anomaly detection

- Simple method for OOD and anomaly detection.
- Qualitative performance matches the state of the art.
- Very low complexity:
  - Deterministic training requiring only PCA in linear case (i.e. matrix SVD).
  - Requires only a few dot-product operations at inference.
  - No re-training of the network parameters (avoids expensive stochastic gradient descent).

Future work

- Extension to visual anomaly segmentation.