Conditional Variational Capsule Network for Open Set Recognition

Yunrui Guo*[1,2] Guglielmo Camporese*[2] Wenjing Yang[1] Alessandro Sperduti[2] Lamberto Ballan[2]

[1]National University of Defense Technology [2]University of Padova

Visual Intelligence Machine Perception Group

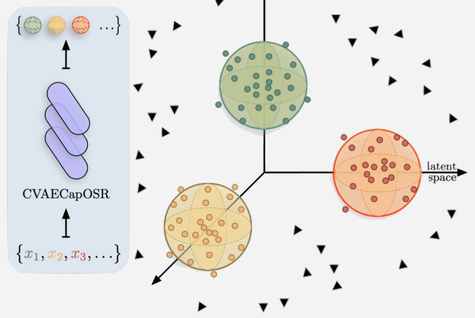
Università degli Studi di Padova

Paper Code

2021 ICCV OCTOBER 11-17 VIRTUAL

Paper: https://arxiv.org/abs/2104.09159
Code: https://github.com/guglielmocamporese/cvaecaposr
* Indicates equal contribution

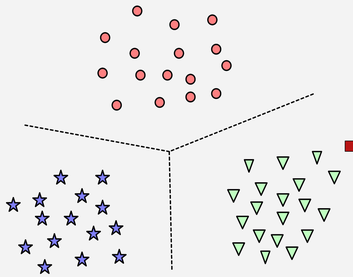
Contributions



- We presented a novel framework for open set recognition based on **CapsNet** and **VAEs**;
- Our model exploits **multiple priors** and it forces the latent representation to follow the gaussian prior selected by the class of the input sample;
- We obtain very competitive results, often outperforming previous state of the art methods by a large margin.

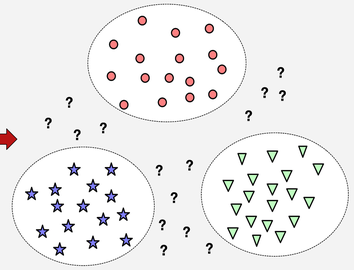
From Closed to Open Set Classification

Classification



- The model has to learn to classify input samples of different classes.

Open Set Classification



- The model has to learn to recognize samples of known classes;
- The model has to detect samples from unknown classes.

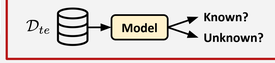
Unknown Detection and Open Set Recognition Problems

Train Stage

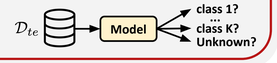
Classification Task
→ Train on *closed multi-class classification.*

Test Stage

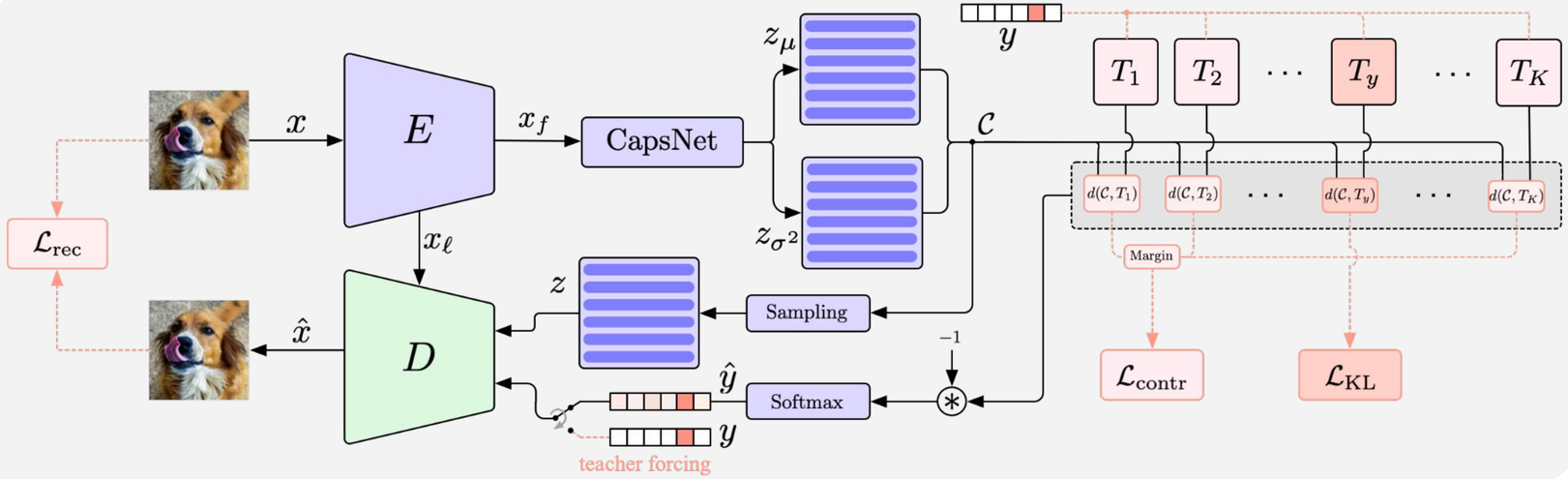
Unknown Detection Task
→ *Detect if a new sample belongs to a known or an unknown class.*



Open Set Recognition Task
→ *Create a new "unknown" class in the classification scenario.*



Model Architecture



Model Building Blocks

Encoding Stage
The input image is fed to a CNN encoder backbone E and then a capsule network produces the probabilistic capsules $\mathcal{C}$ with mean z_μ and z_σ.

Contrastive Variational Stage
In order to map input instances from the same class into compacted and separated regions of the latent space, during the learning process, we let the probabilistic capsule $\mathcal{C}$ to be attracted by the y-th target distribution Ty and at the same time we let all the other targets $\mathbf{T}_{\neq y}$ to be repulsed by $\mathcal{C}$. Using this contrastive strategy, we encourage the encoded representation to belong to the correct region of the latent space while maintaining all the targets sufficiently far apart to each other.

Decoding Stage
The decoder is a convolutional network with transposed convolutions with skip connections. The decoder processes $\mathcal{Z}$ that is a sample from the probabilistic capsule distribution $\mathcal{C}$ added to a learnable embedding that corresponds to the predicted class.

Experimental Results

Unknown Detection Results

Method	MNIST	SVHN	CIFAR10	CIFAR+10	CIFAR+50	TinyImageNet
Softmax †	0.978	0.886	0.677	0.816	0.805	0.577
Openmax †	0.981	0.894	0.695	0.817	0.796	0.576
G-Openmax †	0.984	0.896	0.675	0.827	0.819	0.580
OSRCI †	0.988 ±0.004	0.91 ±0.01	0.699 ±0.038	0.838	0.827	0.586
CROSR	0.991 ±0.004	0.899 ±0.018	-	-	-	0.589
C2AE ‡	-	0.892 ±0.013	0.711 ±0.008	0.810 ±0.005	0.803 ±0.000	0.581 ±0.019
GFROR ‡	-	0.955 ±0.018	0.831 ±0.039	-	-	0.657 ±0.012
CGDL §	0.977 ±0.008	0.896 ±0.023	0.681 ±0.029	0.794 ±0.013	0.794 ±0.003	0.653 ±0.002
RPL §	0.917 ±0.006	0.931 ±0.014	0.784 ±0.025	0.885 ±0.019	0.881 ±0.014	0.711 ±0.026
CVAECapOSR (ours)	**0.992** ±0.004	**0.956** ±0.012	**0.835** ±0.023	**0.888** ±0.019	**0.889** ±0.017	**0.715** ±0.018

AUROC values on the unknown detection task.
For more details see the paper!

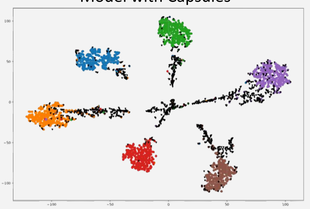
Open Set Recognition Results

Method	ImageNet-crop	ImageNet-resize	LSUN-crop	LSUN-resize
Softmax †	0.639	0.653	0.642	0.647
Openmax †	0.660	0.684	0.657	0.668
CROSR	0.721	0.735	0.720	0.749
C2AE ‡	0.837	0.826	0.783	0.801
CGDL §	0.840	0.832	0.806	0.812
RPL §	0.811	0.810	0.846	0.820
CVAECapOSR (ours)	**0.857**	**0.834**	**0.868**	**0.882**

F1-scores on the open set recognition task.
For more details see the paper!

Known vs Unknown Features Visualization

Model with Capsules



Model without Capsules