FEVER-OOD: Free Energy Vulnerability Elimination for Robust Out-of-Distribution Detection

Supplementary Material

8. Mathematical Background

In this section we provide a detailed analysis of the mathematical formulation of our methods, as described in Sec. 4. Specifically, we analyze the constraints in Eqs. (10) and (11) and the solution to the minimization problem in Eq. (12).

8.1. Null Space Conditions

The singular values of a matrix $A \in \mathbb{R}^{m\times n}$ are defined as the square roots of the eigenvalues of the symmetric matrix $A^\top A \in \mathbb{R}^{n\times n}$. Alternatively, the singular vectors of A can be defined as follows:

$$\mathbf{v}_1 = \underset{\|\mathbf{v}\|_2=1}{\operatorname{argmax}} \|A\mathbf{v}\|_2 \,, \tag{15}$$

$$\mathbf{v}_i = \underset{\substack{\|\mathbf{v}\|_2=1 \\ \mathbf{v}\perp\operatorname{span}\{\mathbf{v}_1,\ldots,\mathbf{v}_{i-1}\}}}{\operatorname{argmax}} \|A\mathbf{v}\|_2, \quad i \geq 2\,. \tag{16}$$

The singular values of A are then given by:

$$\sigma_i(A) = \|A\mathbf{v}_i\|_2\,. \tag{17}$$

The null space of A, denoted as $\operatorname{Null}(A)$, is defined as the span of unit vectors whose corresponding singular values are zero. From Eq. (16), there are n singular vectors (as there are n orthogonal vectors in $\mathbb{R}^n$). If $r = \operatorname{rank}(A)$, the rank-nullity theorem implies that the nullity, *i.e.*, the dimension of $\operatorname{Null}(A)$, is $n-r$. Therefore, from Eq. (16), the null space of A is given by:

$$\operatorname{Null}(A) = \operatorname{span}\{\mathbf{v}_{r+1}, ..., \mathbf{v}_n\}\,. \tag{18}$$

Moreover, if $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are the singular vectors corresponding to the non-zero singular values of A, $\sigma_1(A) \geq \cdots \geq \sigma_r(A) > 0$ (assuming that $A \neq 0$), then the span of these vectors is orthogonal to the null space of A. Formally:

$$\operatorname{Null}(A)^\perp = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}\,. \tag{19}$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are an orthonormal basis of $\mathbb{R}^n$, and the subspaces in Eqs. (18) and (19) are complementary, every feature vector $\boldsymbol{\nu} \in \mathbb{R}^n$ can be written as:

$$\boldsymbol{\nu} = \boldsymbol{\nu}_n + \boldsymbol{\nu}_a\,,\ \boldsymbol{\nu}_n \in \operatorname{Null}(A)\,,\ \boldsymbol{\nu}_a \in \operatorname{Null}(A)^\perp\,. \tag{20}$$

This decomposition indicates that every vector $\boldsymbol{\nu} \in \mathbb{R}^n$ consists of a component $\boldsymbol{\nu}_n$ in the null space of A and a component $\boldsymbol{\nu}_a$ orthogonal to $\operatorname{Null}(A)$. This makes the condition $\boldsymbol{\delta} \in \operatorname{Null}(\mathrm{W}_{cls}^\top)^\perp$ in Eqs. (10) to (12) necessary to avoid a trivial solution to the minimization problem. For instance, if this condition were not enforced, then the solution would be zero, corresponding to any $\boldsymbol{\delta} \in \operatorname{Null}(\mathrm{W}_{cls}^\top)$. If we only enforced $\boldsymbol{\delta} \notin \operatorname{Null}(\mathrm{W}_{cls}^\top)$, then the minimum would not exist but the infimum would be zero:

$$\inf_{\substack{\boldsymbol{\delta}:\|\boldsymbol{\delta}\|\geq d_b, \\ \boldsymbol{\delta}\notin\operatorname{Null}(\mathrm{W}_{cls}^\top)}} \|\mathrm{W}_{cls}^\top\boldsymbol{\delta}\| = 0\,. \tag{21}$$

Eq. (21) holds because $\boldsymbol{\delta}$ can be decomposed into components $\boldsymbol{\delta}_n \in \operatorname{Null}(A)$ and $\boldsymbol{\delta}_a \in \operatorname{Null}(A)^\perp$, where $\boldsymbol{\delta}_n$ can be arbitrarily larger than $\boldsymbol{\delta}_a$, making $\|\mathrm{W}_{cls}^\top\boldsymbol{\delta}\|$ arbitrarily small (but not zero).

8.2. Least Singular Value Solution

From Eqs. (16) and (19), it follows that:

$$\min_{\substack{\|\mathbf{v}\|_2=1 \\ \mathbf{v}\in\operatorname{Null}(A)^\perp}} \|A\mathbf{v}\|_2 = \sigma_r(A) =: \sigma_{min}(A). \tag{22}$$

This equation highlights that the smallest non-zero singular value of A corresponds to the minimum norm of $A\mathbf{v}$ over all unit vectors orthogonal to $\operatorname{Null}(A)$.

The singular value decomposition (SVD) encodes information about the singular values and singular vectors of a given matrix in a structured way. Namely, any matrix $A \in \mathbb{R}^{m\times n}$ can be factorized as follows:

$$A = U\Sigma V^\top\,, \tag{23}$$

where:

- $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal rectangular matrix whose entries are the singular values of A in descending order;
- $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices, whose columns correspond to the left and right singular vectors of A, respectively.

In particular, the SVD of $A^\top$ is given by:

$$A^\top = V\Sigma^\top U^\top\,, \tag{24}$$

which implies that A and $A^\top$ have the same non-zero singular values. Combining this with Equation (22), we get

$$\min_{\substack{\|\mathbf{v}\|_2=1 \\ \mathbf{v}\in\operatorname{Null}(A^\top)^\perp}} \|A^\top\mathbf{v}\|_2 = \sigma_{min}(A)\,, \tag{25}$$

which justifies Equation (12).

8.3. Singular Value Gradients

Lewis and Sendov [26] demonstrate that for a convex and absolutely symmetric function $f : \mathbb{R} \rightarrow (-\infty, +\infty]$, the gradient of the corresponding function $f \circ \sigma$ is differentiable at the matrix X if and only if f is differentiable at $\sigma(X)$, where $\sigma(X)$ are the singular values of X. The gradient is given by:

$$\nabla(f \circ \sigma)(X) = U\text{Diag}\left(\nabla f\left(\sigma\left(X\right)\right)\right)V^\top , \tag{26}$$

where $X = U\text{Diag}\left(X\right)V^\top$. If $(f_i \circ \sigma)(C) = \mathbf{s}_i^\top \sigma(X)$ is a function that selects the i-th singular value, such that $(\mathbf{s}_i)_k = \delta_{ik}$, where δ_{ik} is the Kronecker delta, then $\nabla f\left(\sigma\left(X\right)\right) = \mathbf{s}_i$. Therefore, the gradient of the i-th singular value is:

$$\nabla\sigma_i(X) = \mathbf{u}_i\mathbf{v}_i^\top , \tag{27}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors corresponding to the i-th singular value $\sigma_i(X)$. Furthermore, if the singular values are ordered, then $\sigma_{max} = \sigma(X)_1$ and $\sigma_{min} = \sigma(X)_r$, where r is the rank of X, hence:

$$\nabla\sigma_{min}(X) = \mathbf{u}_{min}\mathbf{v}_{min}^\top \tag{28}$$

$$\nabla\sigma_{max}(X) = \mathbf{u}_{max}\mathbf{v}_{max}^\top . \tag{29}$$

Combining Eq. (28) into the LSV and CN regularizers (Eqs. (13) and (14)), we obtain the gradients:

$$\nabla_{W_{cls}}(\sigma_{min}^{-1}) = -\sigma_{min}^{-2}\mathbf{u}_{min}\mathbf{v}_{min}^\top , \tag{30}$$

$$\nabla_{W_{cls}}(\kappa) = (\sigma_{min}\mathbf{u}_{max}\mathbf{v}_{max}^\top - \sigma_{max}\mathbf{u}_{min}\mathbf{v}_{min}^\top)/\sigma_{min}^2 . \tag{31}$$

9. Training Regime

We follow the original training regime for each baseline method and their corresponding FEVER-OOD variants.

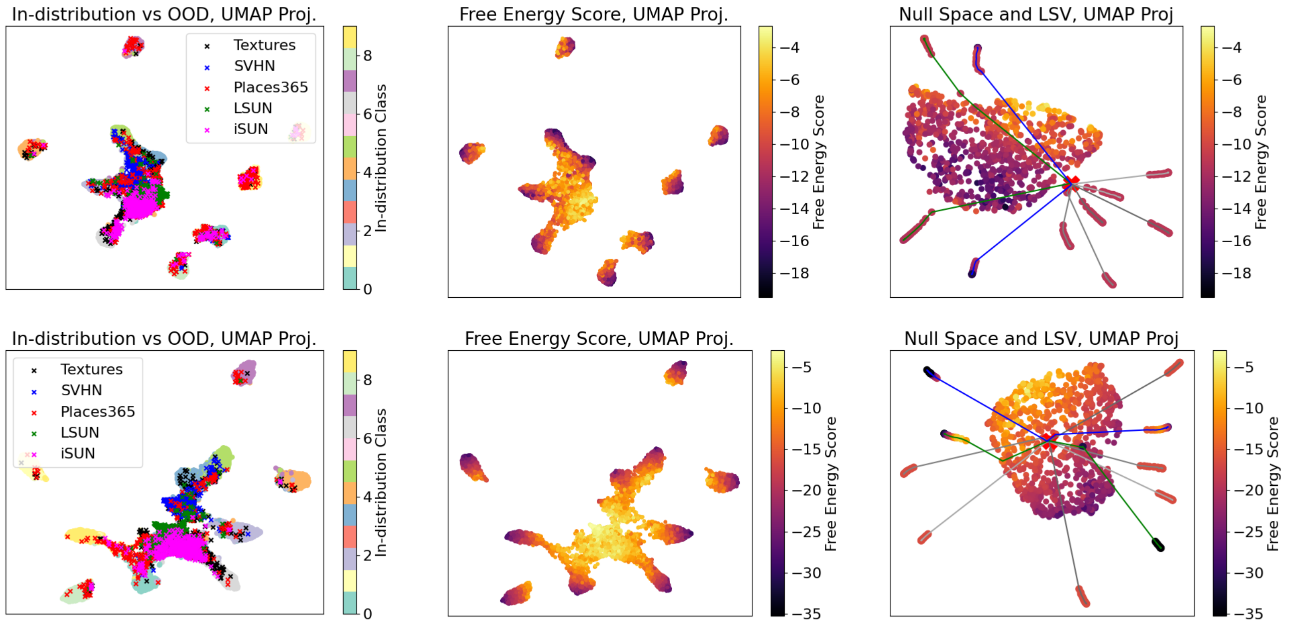
VOS [7]: we train all our VOS for classification models for 100 epochs with a batch size of 128 32 × 32 images (CIFAR-10 and CIFAR-100 [20] datasets). Outlier synthesis started in epoch 40 in all experiments. Following Du *et al.* [7], we sample $10,000$ instances per category in the feature space and choose the instance with the least log probability as the outlier. We use an initial learning rate of 0.1 with cosine annealing and stochastic gradient descent (SGD) with 5×10^{-4} weight decay and 0.9 momentum for all experiments. A loss weight of 0.1 is used for the uncertainty loss. With regards to VOS for detection models, we follow the same outlier synthesis scheme as in classification. We use a batch size of 16 images with varying minimum width from 480 to 800 pixels, and train for $18,000$ iterations, corresponding to around 17.4 epochs for the PASCAL VOC [10] dataset. An initial learning rate of 0.02 is used for all VOS detection models, decaying by a factor of 10 after 12,000 epochs and again after 16,000 epochs. Similarly to classification, loss weight of 0.1 us used for the uncertainty loss. All VOS training was carried out using a single GPU per experiment. VOS classification models were trained on NVIDIA GeForce RTX 2080 Ti GPUs, while VOS detection models were trained on NVIDIA RTX A6000 GPUs.

FFS [21]: FFS models follow a similar training regime as VOS models. The only difference for FFS models is that the outliers are obtained as the least likely out of 200 samples from the normalizing flow feature space, following Kumar *et al.* [21]. We use the same 0.1 loss weight for the uncertainty loss as in VOS, and 1×10^{-4} loss weight for the normalizing flow loss, for both classification and detection models. Additionally, we implement FFS for classification since the original implementation is only for object classification.

Dream-OOD [8]: we train Dream-OOD in CIFAR-100 for 100 epochs with a batch size of 160 in-distribution images and 160 OOD images (for a combined 320 epochs per batch). We use SGD with an initial learning rate of 0.1, cosine annealing, 5×10^{-4} weight decay and 0.9 momentum. Regarding Imagenet-100, we use a ResNet-34 [14] that is pretrained solely on image classification only for 100 epochs. We train for OOD detection for 20 further epochs, using a batch size of 20 in-distribution and 20 OOD images, initial learning rate of 0.001, 5×10^{-4} weight decay and 0.9 momentum. Following Du *et al.* [8], we use an energy loss weight of 2.5 for CIFAR-100 experiments and 1.0 for Imagenet-100 experiments. Dream-OOD for CIFAR-10 models were trained with single NVIDIA GeForce RTX 2080 Ti GPU while the Imagenet-100 models were trained with single NVIDIA TESLA v100 GPUs. We use the ResNet-34 pretrained model for Imagenet-100 and the generated outliers in the pixel space for both CIfAR-100 and Imagenet-100 provided by Du *et al.* [8] at `https://github.com/deeplearning-wisc/dream-ood`.

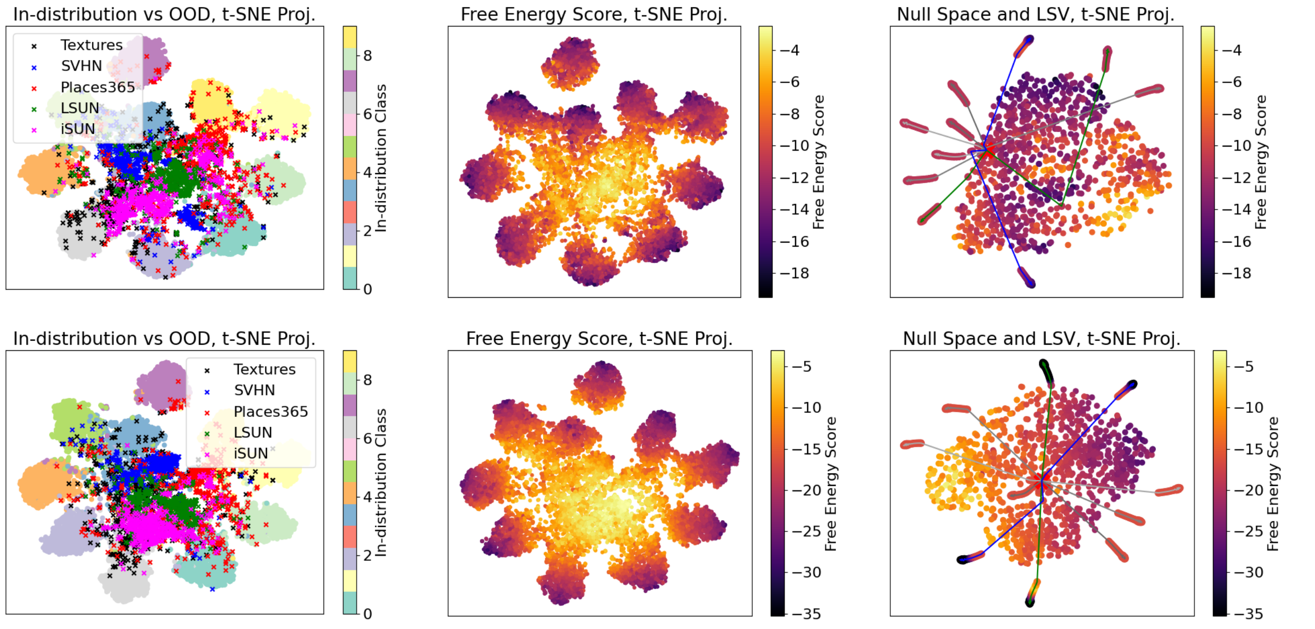
10. Null Space Projection

Figs. 6 and 7 show the feature projections of VOS and FEVER-OOD VOS models using UMAP and t-SNE projections, respectively. Both models are for CIFAR-10 as in-distribution data, with the best model of FEVER-OOD VOS being shown, corresponding to an 96-NSR and $\lambda_{LSV} = 1.0$ (Tab. 1). Specifically, Figs. 6a and 7a show the projection of in-distribution vs. OOD examples. The projections of the feature space of both methods show that the OOD samples are pushed to different regions outside, with defined clusters for in-distribution classes (colored according to their ground-truth class). Nonetheless, the free energy



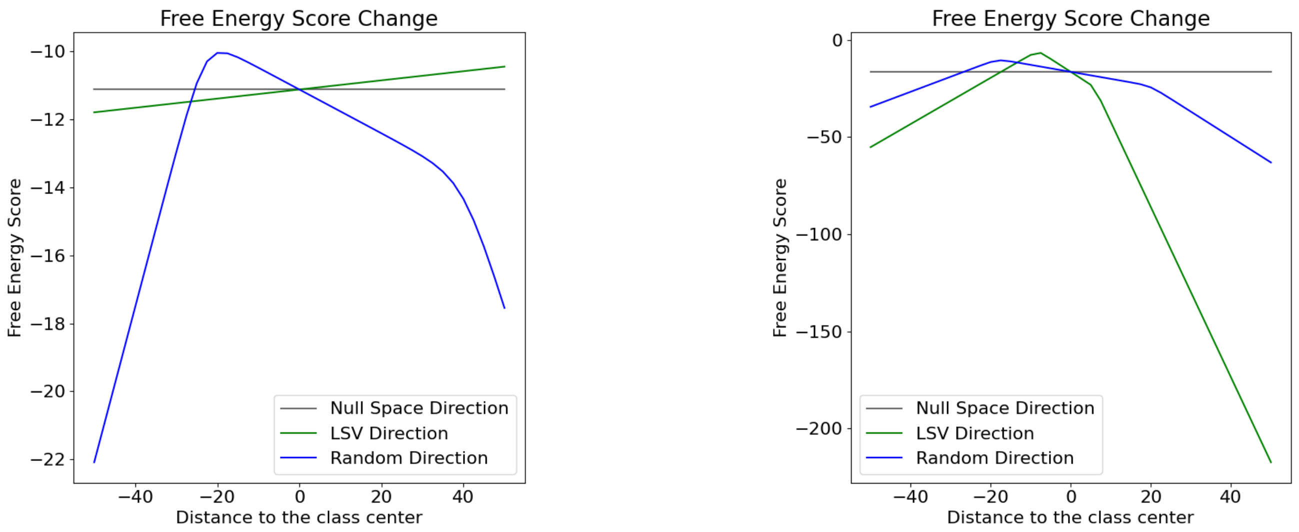
(a) (b) (c)

Figure 6. Feature Space UMAP Projection for models trained on CIFAR-10. Top row corresponds to the VOS [7] model while the bottom shows the FEVER-OOD VOS (Ours) projections. (a) In-distribution vs OOD feature space projection, where × markers represent data from OOD datasets, (b) Free Energy visualization of the feature space, and (c) different important directions, including Null Space directions, the LSV direction and a random direction.



(a) (b) (c)

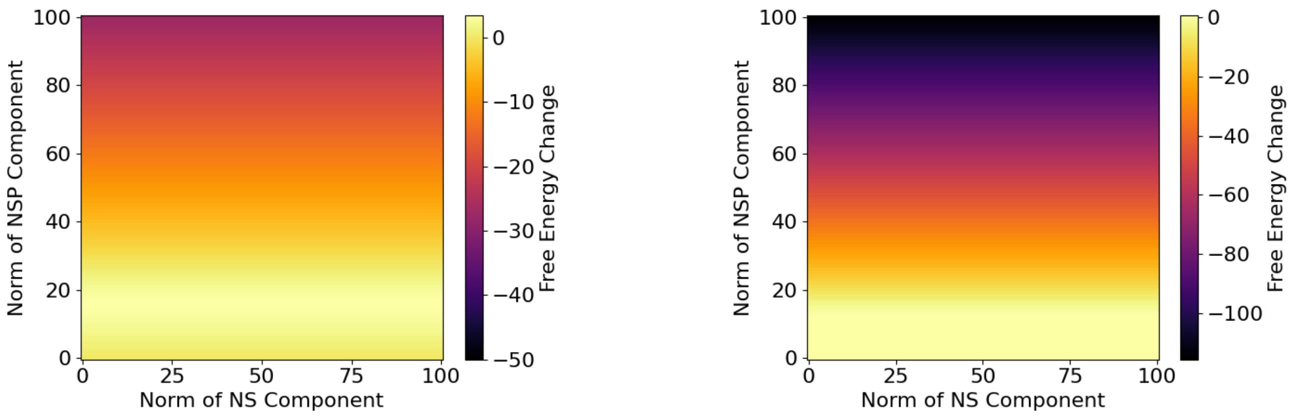
Figure 7. Feature Space t-SNE Projection for models trained on CIFAR-10. Top row corresponds to the VOS [7] model while the bottom shows the FEVER-OOD VOS (Ours) projections. (a) In-distribution vs OOD feature space projection, where × markers represent data from OOD datasets, (b) Free Energy visualization of the feature space, and (c) different important directions, including Null Space directions, the LSV direction and a random direction.



(a) VOS

(b) FEVER-VOS

Figure 8. Free energy change by its distance to the centroid of the feature vectors of an in-distribution category along different directions. (a) VOS vs. (b) FEVER-OOD VOS.



(a) VOS

(b) FEVER-VOS

Figure 9. Free energy change for varying the contribution of the Null Space (NS) component and the Null Space Perpendicular (NSP) component. (a) VOS vs. (b) FEVER-OOD VOS.

score for these samples shown in Figs. 6b and 7b exhibits a more uniform distribution of the free energy for OOD samples when using FEVER-OOD, providing a visualization of why our technique works better. Additionally, the different scales of the energy values between both models indicates a higher separability between in-distribution and OOD. Finally, Figs. 6c and 7c show the feature distribution of the class 0 and some virtually generated features along some directions. Specifically, we generate features away from the centroid of the in-distribution feature vectors along three null space directions (gray), the LSV direction (green) and a random direction (blue). As described in Sec. 4.1, the energy of the features in the null space direction does not change, which is reflected in Figs. 6c and 7c, where all the null space features have the same energy. Additionally, Sec. 4.2 shows that the direction with least change in energy corresponds to the LSV direction. In this sense, because our FEVER-OOD VOS (Figs. 6c and 7c) uses the LSVR, the energy plots show greater energy variation in the LSV direction (green line) compared to the baseline VOS.

Fig. 8 shows the change in energy across these directions with respect to the distance to the in-distribution centre, where it is seen that the change in energy in the LSV direction is significantly larger with FEVER-OOD. Finally, with

Table 5. CIFAR-10 Results (ID Acc. = in-distribution accuracy; Null Space Reduction (NSR) methods = our approach).

Method	FEVER-OOD			OOD Datasets												ID Acc
				Textures		SHVN		Places365		LSUN		iSUN		Avg		
	r-NSR	λ_{LSV}	λ_{CN}	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	
VOS	-	-	-	50.05$_{\pm 5.70}$	86.79$_{\pm 1.28}$	39.15$_{\pm 10.52}$	91.73$_{\pm 2.88}$	40.88$_{\pm 1.75}$	89.54$_{\pm 0.69}$	8.22$_{\pm 1.02}$	98.36$_{\pm 0.20}$	30.01$_{\pm 4.44}$	94.31$_{\pm 0.84}$	33.66$_{\pm 2.21}$	92.15$_{\pm 0.59}$	94.83$_{\pm 0.16}$
	-	1.0	-	55.58$_{\pm 6.35}$	86.71$_{\pm 2.35}$	43.00$_{\pm 3.16}$	92.09$_{\pm 0.94}$	38.78$_{\pm 2.51}$	91.14$_{\pm 0.56}$	7.36$_{\pm 1.59}$	98.46$_{\pm 0.32}$	21.23$_{\pm 6.95}$	96.23$_{\pm 1.17}$	33.19$_{\pm 3.20}$	92.92$_{\pm 0.85}$	94.66$_{\pm 0.18}$
	-	-	0.1	50.80$_{\pm 3.06}$	85.39$_{\pm 0.79}$	30.78$_{\pm 2.95}$	94.04$_{\pm 0.58}$	41.60$_{\pm 1.40}$	88.77$_{\pm 0.74}$	8.50$_{\pm 0.82}$	98.30$_{\pm 0.10}$	30.91$_{\pm 4.54}$	93.41$_{\pm 1.68}$	32.52$_{\pm 1.25}$	91.98$_{\pm 0.46}$	94.69$_{\pm 0.16}$
	96	-	-	50.27$_{\pm 2.32}$	89.42$_{\pm 0.49}$	36.25$_{\pm 14.04}$	93.93$_{\pm 2.00}$	39.51$_{\pm 1.68}$	91.42$_{\pm 0.52}$	6.98$_{\pm 0.91}$	98.63$_{\pm 0.18}$	22.95$_{\pm 7.32}$	96.02$_{\pm 1.32}$	31.19$_{\pm 3.21}$	93.89$_{\pm 0.47}$	94.69$_{\pm 0.16}$
	96	1.0	-	**40.86**$_{\pm 4.83}$	**92.12**$_{\pm 1.05}$	41.92$_{\pm 16.11}$	93.20$_{\pm 2.61}$	**35.69**$_{\pm 2.29}$	**92.61**$_{\pm 0.57}$	**4.95**$_{\pm 1.49}$	**98.93**$_{\pm 0.24}$	**17.68**$_{\pm 5.26}$	**97.04**$_{\pm 0.85}$	**28.22**$_{\pm 4.18}$	**94.78**$_{\pm 0.85}$	94.74$_{\pm 0.17}$
	96	-	0.1	43.60$_{\pm 3.32}$	90.86$_{\pm 0.71}$	48.71$_{\pm 11.23}$	91.99$_{\pm 1.95}$	38.47$_{\pm 1.36}$	91.34$_{\pm 0.54}$	5.90$_{\pm 1.66}$	98.76$_{\pm 0.26}$	18.27$_{\pm 5.40}$	96.81$_{\pm 0.83}$	30.99$_{\pm 3.47}$	93.95$_{\pm 0.56}$	94.65$_{\pm 0.12}$
	64	-	-	46.97$_{\pm 1.35}$	89.57$_{\pm 0.64}$	37.79$_{\pm 6.83}$	93.45$_{\pm 1.00}$	36.70$_{\pm 1.51}$	91.84$_{\pm 0.50}$	6.60$_{\pm 0.91}$	98.66$_{\pm 0.14}$	22.29$_{\pm 3.18}$	96.00$_{\pm 0.68}$	30.07$_{\pm 1.71}$	93.91$_{\pm 0.38}$	94.76$_{\pm 0.07}$
	64	1.0	-	44.41$_{\pm 5.39}$	91.26$_{\pm 1.29}$	34.68$_{\pm 10.93}$	94.39$_{\pm 1.72}$	35.85$_{\pm 2.56}$	92.42$_{\pm 0.53}$	5.68$_{\pm 0.57}$	98.75$_{\pm 0.13}$	24.03$_{\pm 5.91}$	96.02$_{\pm 0.95}$	28.93$_{\pm 3.86}$	94.57$_{\pm 0.66}$	94.75$_{\pm 0.13}$
	64	-	0.001	48.54$_{\pm 4.50}$	89.40$_{\pm 1.38}$	45.51$_{\pm 9.64}$	91.41$_{\pm 2.42}$	38.34$_{\pm 1.74}$	91.39$_{\pm 0.49}$	6.99$_{\pm 0.87}$	98.64$_{\pm 0.13}$	21.29$_{\pm 4.28}$	96.28$_{\pm 0.72}$	32.13$_{\pm 3.01}$	93.42$_{\pm 0.74}$	94.78$_{\pm 0.10}$
	32	-	-	43.89$_{\pm 3.14}$	90.66$_{\pm 0.81}$	50.86$_{\pm 10.43}$	91.48$_{\pm 1.74}$	37.95$_{\pm 0.95}$	91.73$_{\pm 0.43}$	5.60$_{\pm 0.74}$	98.84$_{\pm 0.12}$	24.57$_{\pm 6.02}$	95.55$_{\pm 1.33}$	32.57$_{\pm 2.09}$	93.65$_{\pm 0.60}$	94.75$_{\pm 0.14}$
	32	0.001	-	46.53$_{\pm 3.77}$	89.67$_{\pm 1.09}$	27.84$_{\pm 9.47}$	95.29$_{\pm 1.08}$	37.33$_{\pm 1.63}$	91.77$_{\pm 0.39}$	5.65$_{\pm 0.77}$	98.82$_{\pm 0.11}$	24.09$_{\pm 6.77}$	95.68$_{\pm 1.35}$	28.29$_{\pm 3.33}$	94.25$_{\pm 0.55}$	94.68$_{\pm 0.07}$
	32	-	0.1	48.25$_{\pm 5.47}$	89.32$_{\pm 1.39}$	**25.31**$_{\pm 2.15}$	**95.71**$_{\pm 0.41}$	39.49$_{\pm 2.75}$	91.23$_{\pm 0.83}$	7.09$_{\pm 1.37}$	98.63$_{\pm 0.24}$	26.50$_{\pm 7.71}$	95.00$_{\pm 2.19}$	29.33$_{\pm 1.29}$	93.98$_{\pm 0.44}$	94.84$_{\pm 0.17}$
	10	-	-	53.20$_{\pm 3.90}$	88.97$_{\pm 0.73}$	35.62$_{\pm 9.66}$	94.41$_{\pm 1.28}$	45.73$_{\pm 6.29}$	89.73$_{\pm 1.69}$	11.72$_{\pm 3.39}$	97.93$_{\pm 0.46}$	41.80$_{\pm 14.35}$	92.51$_{\pm 3.03}$	37.61$_{\pm 4.00}$	92.71$_{\pm 0.97}$	91.95$_{\pm 2.32}$
	10	0.01	-	72.61$_{\pm 22.80}$	72.98$_{\pm 18.81}$	76.83$_{\pm 24.26}$	73.13$_{\pm 19.16}$	72.09$_{\pm 24.70}$	70.86$_{\pm 17.48}$	49.57$_{\pm 41.32}$	78.19$_{\pm 23.02}$	59.87$_{\pm 32.88}$	76.66$_{\pm 21.76}$	66.19$_{\pm 28.10}$	74.37$_{\pm 19.94}$	58.10$_{\pm 39.34}$
	10	-	0.001	67.02$_{\pm 17.39}$	80.09$_{\pm 15.10}$	58.11$_{\pm 24.43}$	83.56$_{\pm 16.90}$	58.03$_{\pm 21.22}$	81.80$_{\pm 15.94}$	34.06$_{\pm 33.25}$	87.46$_{\pm 18.75}$	56.56$_{\pm 24.29}$	83.65$_{\pm 17.01}$	54.76$_{\pm 22.82}$	83.31$_{\pm 16.67}$	74.13$_{\pm 32.08}$
FFS	-	-	-	52.86$_{\pm 4.49}$	83.47$_{\pm 1.67}$	38.67$_{\pm 11.21}$	89.74$_{\pm 5.19}$	44.65$_{\pm 1.29}$	87.47$_{\pm 0.77}$	6.59$_{\pm 0.92}$	98.67$_{\pm 0.17}$	31.34$_{\pm 1.88}$	93.24$_{\pm 0.97}$	34.82$_{\pm 2.03}$	90.52$_{\pm 0.98}$	94.69$_{\pm 0.15}$
	-	0.001		50.74$_{\pm 3.98}$	84.93$_{\pm 0.75}$	32.25$_{\pm 12.16}$	92.97$_{\pm 3.60}$	42.99$_{\pm 2.30}$	88.30$_{\pm 1.10}$	5.76$_{\pm 1.23}$	98.82$_{\pm 0.27}$	28.16$_{\pm 6.45}$	94.18$_{\pm 1.98}$	31.98$_{\pm 3.10}$	91.84$_{\pm 1.28}$	94.73$_{\pm 0.12}$
	-	-	1.0	50.08$_{\pm 3.86}$	84.92$_{\pm 1.85}$	30.91$_{\pm 6.19}$	92.77$_{\pm 2.33}$	45.69$_{\pm 2.59}$	87.16$_{\pm 0.75}$	6.22$_{\pm 0.65}$	98.73$_{\pm 0.11}$	26.60$_{\pm 3.64}$	94.80$_{\pm 0.64}$	31.90$_{\pm 2.51}$	91.67$_{\pm 0.94}$	94.85$_{\pm 0.20}$
	96	-	-	48.67$_{\pm 2.88}$	88.70$_{\pm 0.48}$	40.64$_{\pm 19.36}$	89.89$_{\pm 9.15}$	41.38$_{\pm 1.20}$	90.22$_{\pm 0.88}$	4.84$_{\pm 0.81}$	99.02$_{\pm 0.13}$	27.30$_{\pm 6.08}$	94.99$_{\pm 1.18}$	32.57$_{\pm 4.52}$	92.56$_{\pm 2.03}$	94.71$_{\pm 0.13}$
	96	1.0	-	48.11$_{\pm 5.15}$	89.84$_{\pm 1.51}$	40.65$_{\pm 15.47}$	93.32$_{\pm 3.16}$	**36.37**$_{\pm 2.28}$	**92.29**$_{\pm 0.60}$	5.22$_{\pm 1.35}$	98.89$_{\pm 0.18}$	24.84$_{\pm 10.08}$	95.82$_{\pm 1.68}$	31.04$_{\pm 5.05}$	**94.03**$_{\pm 1.06}$	94.71$_{\pm 0.08}$
	96	-	0.001	47.24$_{\pm 2.61}$	89.51$_{\pm 0.88}$	39.25$_{\pm 10.20}$	93.84$_{\pm 1.33}$	40.44$_{\pm 3.32}$	90.83$_{\pm 0.97}$	**4.44**$_{\pm 0.81}$	**99.08**$_{\pm 0.12}$	25.27$_{\pm 6.08}$	95.71$_{\pm 0.98}$	31.33$_{\pm 3.13}$	93.80$_{\pm 0.60}$	94.84$_{\pm 0.14}$
	64	-	-	**45.03**$_{\pm 2.88}$	**89.96**$_{\pm 1.01}$	37.96$_{\pm 2.68}$	93.18$_{\pm 1.53}$	41.70$_{\pm 1.51}$	90.49$_{\pm 0.57}$	4.64$_{\pm 0.30}$	99.03$_{\pm 0.04}$	**22.78**$_{\pm 5.67}$	**96.02**$_{\pm 0.97}$	30.42$_{\pm 2.09}$	93.73$_{\pm 0.69}$	94.82$_{\pm 0.07}$
	64	1.0	-	51.70$_{\pm 5.67}$	88.81$_{\pm 1.09}$	33.02$_{\pm 14.09}$	94.60$_{\pm 2.30}$	39.05$_{\pm 3.73}$	91.53$_{\pm 1.23}$	5.56$_{\pm 1.95}$	98.84$_{\pm 0.30}$	26.63$_{\pm 7.51}$	95.65$_{\pm 1.12}$	31.19$_{\pm 4.32}$	93.89$_{\pm 0.69}$	94.68$_{\pm 0.08}$
	64	-	0.001	47.16$_{\pm 6.00}$	88.88$_{\pm 2.08}$	**25.18**$_{\pm 9.52}$	**95.65**$_{\pm 1.62}$	43.52$_{\pm 2.75}$	89.96$_{\pm 0.67}$	5.01$_{\pm 0.84}$	99.03$_{\pm 0.13}$	28.45$_{\pm 8.77}$	94.79$_{\pm 1.71}$	**29.86**$_{\pm 3.34}$	93.66$_{\pm 0.79}$	94.79$_{\pm 0.19}$
	32	-	-	46.75$_{\pm 3.49}$	89.61$_{\pm 1.40}$	31.77$_{\pm 8.73}$	94.44$_{\pm 1.35}$	41.51$_{\pm 3.38}$	90.72$_{\pm 1.02}$	4.93$_{\pm 1.11}$	99.04$_{\pm 0.16}$	28.92$_{\pm 11.23}$	94.49$_{\pm 2.94}$	30.78$_{\pm 3.70}$	93.66$_{\pm 0.95}$	94.79$_{\pm 0.18}$
	32	0.001	-	49.79$_{\pm 3.05}$	88.29$_{\pm 0.93}$	31.60$_{\pm 16.07}$	94.14$_{\pm 3.53}$	41.25$_{\pm 1.84}$	90.58$_{\pm 0.70}$	5.00$_{\pm 1.28}$	98.99$_{\pm 0.19}$	28.24$_{\pm 6.75}$	95.06$_{\pm 1.22}$	31.18$_{\pm 5.21}$	93.41$_{\pm 1.03}$	94.74$_{\pm 0.17}$
	32	-	0.01	51.83$_{\pm 5.63}$	87.07$_{\pm 1.91}$	33.13$_{\pm 6.68}$	94.35$_{\pm 1.40}$	46.81$_{\pm 2.95}$	88.85$_{\pm 1.02}$	5.37$_{\pm 1.64}$	98.95$_{\pm 0.28}$	31.78$_{\pm 8.93}$	94.56$_{\pm 1.52}$	33.78$_{\pm 3.89}$	92.75$_{\pm 1.01}$	94.70$_{\pm 0.19}$
	10	-	-	54.72$_{\pm 5.89}$	84.77$_{\pm 3.23}$	45.70$_{\pm 12.95}$	91.49$_{\pm 2.27}$	41.52$_{\pm 3.27}$	89.51$_{\pm 1.31}$	8.21$_{\pm 3.59}$	98.33$_{\pm 0.67}$	28.58$_{\pm 7.75}$	94.30$_{\pm 1.82}$	35.75$_{\pm 5.26}$	91.68$_{\pm 1.38}$	94.54$_{\pm 0.17}$
	10	0.001	-	62.04$_{\pm 19.73}$	79.12$_{\pm 14.68}$	55.11$_{\pm 27.80}$	81.97$_{\pm 16.88}$	54.23$_{\pm 23.47}$	81.49$_{\pm 15.84}$	26.86$_{\pm 36.64}$	88.65$_{\pm 19.33}$	49.99$_{\pm 28.02}$	84.06$_{\pm 17.28}$	49.65$_{\pm 25.60}$	83.06$_{\pm 16.58}$	77.76$_{\pm 33.88}$
	10	-	0.001	66.08$_{\pm 18.34}$	75.24$_{\pm 13.01}$	62.20$_{\pm 29.49}$	80.75$_{\pm 16.21}$	56.53$_{\pm 22.11}$	79.83$_{\pm 15.02}$	31.00$_{\pm 34.71}$	87.73$_{\pm 18.88}$	51.68$_{\pm 24.86}$	83.40$_{\pm 16.74}$	53.50$_{\pm 24.12}$	81.39$_{\pm 15.82}$	77.62$_{\pm 33.81}$

regards to the component decomposition in Eq. (20), Fig. 9 shows the energy change with respect to an in-distribution feature vector when moving in directions with varying contribution of the components of the null space and perpendicular to the null space. Since the null space component does not change the free energy score, all the changes are in the vertical direction, showing a greater change when using FEVER-OOD vs. the baseline methods.

11. Ablation Studies

Tabs. 5 to 7 show more extensive results of different combinations of NSR, LSVR and CNR in the FEVER-OOD framework. The best values for λ_{LSV} and λ_{CN} are reported. In general, it is observed that using LSVR improves the detection of the baseline and NSR versions, while CNR usually decreases the AUROC. It is also seen that NSR also increases AUROC and decreases FPR95. Nonetheless, excessive NSR in feature-outlier synthesis methods(*e.g.*, VOS using 10-NSR for CIFAR-10) can negatively impact OOD detection performance. On the other hand, Dream-OOD models achieve better results with significant NSR, although LSVR and CNR seem to only be beneficial for Imagenet-100.

Figs. 10 to 14 show the ablation studies of varying λ_{LSV} and λ_{CN} in Eqs. (13) and (14) for different classification methods and in-distribution datasets. Fig. 10 shows the results for FEVER-OOD VOS using CIFAR-10 as in-distribution, where values corresponding to a $\lambda_{\{LSV,CN\}} = 0$ refer to no LSV or CN regularisation. In general, LSV regularisation gives better results than CN regularisation. It is observed that larger values of λ_{LSV} leads to better performance for none or few NSR (VOS, VOS-96-NSR and VOS-64-NSR). The same trend is observed for FEVER-OOD FFS in Fig. 11. It is also observed that the models become unstable when using a large NSR, where the extreme case of {VOS,FFS}-10-NSR fails for both regularizer at relative small loss weights. This effect could be caused because regualrizing the least singular value (either for LSVR or CNR) affects all the directions of the feature space since there is no null space. This causes makes the model not able to learn the in-distribution task, failing also for OOD detection. Additionally, all NSR models fail for $\lambda_{CN} = 1.0$.

FEVER-OOD VOS and FFS ablations for the CIFAR-100 as in-distribution are shown in Figs. 12 and 13. Similar as with CIFAR-10, LSV regularisation is more stable and leads to better results than CN regularisation. In both OOD models (VOS and FFS), it is observed that the best results are achieved with 114-NSR and an intermediate value for λ_{LSV}. These results indicate that for OOD models based in outlier generation in the feature space, some reduction of the null space and a moderate regularizer is benefical. However, the complete elimination of the null space and LSV (or CN) regularization might impose a huge prior in the last layer, making it difficult to learn the in-distribution task. Finally, Fig. 14 shows the ablations for FEVER-OOD Dream-OOD with CIFAR-100 as in-distribution. Here it is observed that NSR by itself leads to better results, suggesting that there might be a significant portion of generated outliers in with large components in the null space

Table 6. CIFAR-100 Results (ID Acc. = in-distribution accuracy; Null Space Reduction (NSR) methods = our approach).

Method	FEVER-OOD			OOD Datasets												ID Acc
				Textures		SHVN		Places365		LSUN		iSUN		Avg		
	r-NSR	λ_{LSV}	λ_{CN}	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	
VOS	-	-	-	81.24±2.11	76.01±0.80	76.09±6.51	83.18±3.59	80.53±1.46	76.25±0.78	37.86±4.06	93.21±0.68	79.34±1.60	76.43±1.44	71.01±1.23	81.02±0.40	76.04±0.22
	-	0.1	-	82.87±1.11	76.49±0.94	**73.35**±2.96	84.57±1.53	80.36±1.16	76.89±0.79	40.05±3.74	92.83±0.46	74.28±4.03	80.25±3.34	70.18±0.87	82.21±0.56	76.01±0.16
	-	-	0.01	82.51±2.64	76.04±1.91	73.67±3.10	**85.31**±1.24	80.18±1.08	76.59±0.23	41.08±3.27	92.75±0.63	80.83±4.37	73.79±5.51	71.65±0.96	80.90±1.11	76.20±0.11
	114	-	-	80.63±0.96	79.06±1.06	81.92±7.49	83.25±4.14	80.23±0.69	77.02±0.34	28.61±1.69	95.20±0.31	76.87±6.24	79.18±3.79	69.65±2.94	82.74±1.65	75.40±0.31
	114	0.01	-	80.02±2.75	78.73±1.83	83.45±2.59	82.07±1.36	78.41±0.96	77.71±0.93	28.90±3.61	95.11±0.58	**72.07**±9.07	80.55±4.70	**68.57**±2.28	**82.83**±0.91	75.72±0.16
	114	-	0.001	**79.37**±1.87	**79.19**±1.03	86.96±1.93	79.25±2.15	**78.20**±0.48	**77.91**±0.40	**26.05**±2.38	**95.70**±0.34	72.95±4.35	**81.12**±2.45	68.71±1.18	82.63±0.83	75.44±0.20
	100	-	-	80.47±2.32	78.21±1.14	79.10±7.42	83.39±3.91	79.19±1.04	77.23±0.50	28.99±2.59	95.15±0.39	75.51±6.68	79.63±3.70	68.65±1.29	82.72±0.57	75.54±0.19
	100	0.001	-	82.60±8.93	72.91±11.54	88.71±7.23	72.19±11.75	82.94±8.56	71.84±10.94	43.40±28.32	86.02±18.01	81.55±9.90	72.74±11.58	75.84±12.12	75.14±12.58	60.57±29.78
	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FFS	-	-	-	82.87±1.39	75.54±0.93	76.32±6.17	84.83±2.36	81.14±1.09	76.27±0.74	36.27±4.91	93.54±1.03	82.83±2.98	74.29±3.08	71.89±1.64	80.89±1.11	76.04±0.11
	-	0.01	-	80.95±1.34	76.53±0.70	83.12±5.59	81.61±2.92	80.19±0.84	76.58±0.49	32.93±4.05	94.20±0.51	79.05±5.16	77.28±3.23	71.25±2.11	81.24±0.72	76.27±0.28
	-	-	0.001	80.27±2.76	76.78±1.53	78.57±7.32	82.57±2.95	80.40±1.09	76.40±0.53	32.91±4.90	94.06±1.02	80.71±7.90	74.49±5.57	70.57±2.41	80.86±0.88	76.05±0.16
	114	-	-	80.99±1.36	78.17±0.83	83.70±7.11	80.78±5.09	80.24±0.79	77.02±0.45	25.36±2.40	95.73±0.40	76.10±6.85	79.87±3.14	69.28±1.95	82.31±1.27	75.60±0.23
	114	0.001	-	79.26±2.67	78.16±1.50	**74.28**±7.13	**85.37**±3.20	**79.45**±0.85	**77.25**±0.61	24.10±1.42	95.94±0.17	**73.61**±5.82	**80.61**±2.49	**66.14**±1.91	**83.47**±0.98	75.45±0.37
	114	-	0.001	81.77±1.81	76.87±1.76	82.61±7.99	78.78±6.21	80.14±1.05	77.08±0.82	26.15±1.91	95.52±0.27	75.69±7.73	78.84±3.69	69.27±1.69	81.42±0.96	75.14±0.25
	100	-	-	**77.69**±2.97	**78.62**±1.02	77.84±9.62	83.40±3.52	80.41±0.72	76.58±0.35	**21.91**±0.95	**96.25**±0.11	79.60±6.06	75.79±5.17	67.49±2.03	82.13±0.85	75.48±0.28
	100	0.001	-	84.73±7.80	70.68±10.41	86.40±7.94	74.77±12.47	83.15±8.50	71.57±10.79	39.78±30.19	86.59±18.30	82.75±8.98	72.08±11.35	75.36±12.36	75.14±12.57	60.52±29.76
	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dream-OOD	-	-	-	62.20±1.02	83.84±0.38	73.05±1.92	84.56±0.21	77.95±1.97	79.43±0.17	39.90±2.01	92.87±0.44	1.70±0.11	99.58±0.04	50.96±1.44	88.06±0.60	75.61±0.19
	-	0.01	-	58.45±2.27	86.04±0.41	68.75 ±1.84	87.65±0.30	77.45±2.11	78.59±0.22	**15.45**±2.57	**97.24**±0.09	1.55±0.03	99.63±0.45	44.33±3.48	89.83±0.12	75.87±0.23
	-	-	0.001	57.4±1.88	86.28±0.32	77.75±1.11	85.13±0.41	78.6±1.23	78.73±0.15	27.2±1.77	95.01±0.42	1.55±0.08	99.57±0.07	48.5±1.54	88.94±0.43	76.32±0.14
	256	-	-	60.00±3.21	85.42±0.79	67.50±2.47	85.84±0.66	75.90±1.10	79.57±0.43	19.85±1.05	96.74±0.67	**1.00**±0.02	**99.78**±0.06	44.85±2.72	89.47±0.242	77.01±0.12
	256	0.001	-	54.25±2.13	86.18±0.86	82.60±2.24	81.70±0.91	**71.20**±1.33	**81.23**±0.61	23.45±1.31	95.87±0.41	1.05±0.02	99.69±0.05	46.51±0.99	88.93±0.14	76.40±0.30
	256	-	1	60.05±2.56	84.52±0.11	61.75±2.03	88.30±0.40	78.15±2.72	76.60±0.82	34.80±1.09	93.18±0.39	1.50±0.14	99.69±0.10	47.25±1.65	88.46±0.68	77.49±0.15
	128	-	-	**52.55**±1.95	87.44±0.24	73.95±1.46	80.05±0.53	71.9±2.48	81.17±0.23	17.7±0.91	97.07±0.13	1.3±0.05	99.73±0.16	43.48±2.22	89.09±0.67	76.72±1.21
	128	0.1	-	56.45±2.39	**87.73**±1.01	67.45±1.12	87.53±0.44	78.45±2.46	77.24±0.61	29.6±1.73	94.81±0.40	1.95±0.10	99.5±0.10	46.78±2.72	89.36±0.83	76.21±0.35
	128	-	0.01	56.05±2.38	86.35±0.79	83.1±1.95	80.75±0.64	75.95±2.53	79.81±0.37	23.5±0.98	95.66±0.30	1.45±0.12	99.65±0.02	48.01±1.54	88.44±0.55	76.35±0.27
	100	-	-	57.55±2.48	86.8±0.44	**54.45**±3.02	**88.69**±0.59	75.8±1.21	78.88±0.69	24.45±1.57	95.86±0.52	1.6±0.07	99.65±0.21	**42.77**±2.10	**89.98**±0.15	76.41±0.32
	100	0.001	-	82.6±4.73	62.38±2.23	90.9±3.33	60.9±1.52	84.4±3.72	69.39±2.27	67.05±1.30	72.8±1.90	5.8±0.85	97.81±0.49	66.15±1.79	72.66±1.17	31.76±4.15
	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 7. ImageNet-100 Results (ID Acc. = in-distribution accuracy; Null Space Reduction (NSR) methods = our approach).

Method	FEVER-OOD			OOD Datasets										ID Acc
				iNaturalist		Places365		SUN		Textures		Avg		
	r-NSR	λ_{LSV}	λ_{CN}	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	
Dream-OOD	-	-	-	23.98±2.08	95.94±0.26	41.75±1.72	92.48±0.12	40.85±1.13	92.76±0.09	50.73±0.85	86.21±0.29	39.33±1.08	91.84±0.12	87.76±0.19
	-	0.01	-	59.91±3.09	90.89±0.48	66.26±2.17	88.12±0.47	74.26±1.52	84.53±0.33	60.56±1.21	83.91±0.42	65.25±1.50	86.87±0.25	87.85±0.11
	-	-	0.001	59.71±3.53	90.97±0.70	65.01±1.90	88.39±0.45	73.93±1.58	84.69±0.49	60.51±1.46	84.01±0.51	64.79±1.69	87.01±0.45	87.77±0.13
	256	-	-	23.84±1.73	95.80±0.30	44.35±1.71	92.14±0.30	42.92±1.98	92.46±0.33	44.38±1.81	88.60±0.54	38.87±1.29	92.25±0.23	87.57±0.24
	256	0.01	-	24.02±1.48	95.75±0.26	44.09±1.87	92.02±0.27	43.36±1.53	92.30±0.30	44.55±1.38	88.67±0.39	39.00±1.05	92.18±0.24	87.54±0.24
	256	-	0.001	24.42±1.43	95.73±0.30	44.98±2.24	91.98±0.36	43.88±1.63	92.21±0.27	44.76±1.16	88.62±0.57	39.51±0.72	92.13±0.18	87.63±0.18
	128	-	-	23.96±2.93	95.91±0.38	43.73±2.50	92.25±0.33	43.44±2.69	92.39±0.41	42.38±1.64	89.40±0.33	38.38±2.21	92.49±0.24	87.53±0.12
	128	0.01	-	23.46±2.54	95.93±0.36	43.37±1.47	92.19±0.22	42.42±1.60	92.44±0.29	**41.86**±1.09	**89.48**±0.27	37.78±1.45	92.51±0.22	87.60±0.11
	128	-	0.001	24.55±2.47	95.87±0.30	44.12±2.03	92.20±0.26	43.89±2.21	92.40±0.33	42.53±1.38	**89.48**±0.40	38.77±1.76	92.49±0.23	87.49±0.11
	100	-	-	23.13±1.35	96.00±0.21	42.49±2.30	92.37±0.37	41.62±2.37	92.69±0.31	41.88±1.11	89.34±0.23	37.28±1.40	92.60±0.18	87.44±0.07
	100	0.01	-	**22.24**±0.81	**96.16**±0.14	**41.08**±3.03	**92.59**±0.44	**40.39**±2.84	**92.91**±0.40	42.31±0.97	89.30±0.37	**36.50**±1.67	**92.74**±0.21	87.42±0.15
	100	-	0.001	23.26±1.71	96.03±0.26	43.09±2.48	92.40±0.29	41.91±3.05	92.71±0.40	43.49±0.58	89.13±0.42	37.94±1.51	92.57±0.17	87.58±0.11

of the feature space. Dream-OOD follows a similar patter as the other models for CIFAR-100, suggesting that the analysis holds for different OOD approaches. As shown in Tabs. 8 to 10, our energy-based method consistently outperforms non-energy-based approaches on most OOD datasets, achieving a lower FPR and a higher AUROC.

12. Qualitative Results

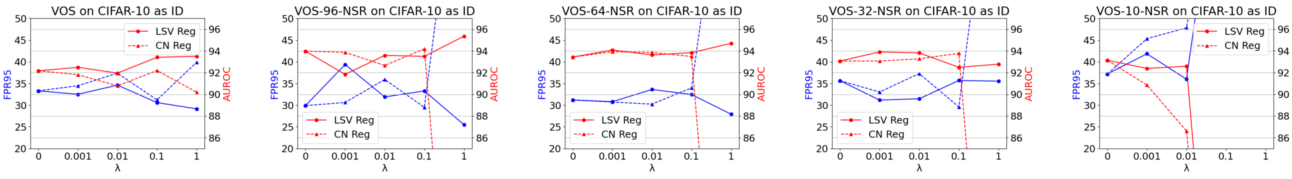
Additional qualitative examples for object-level OOD detection using VOS [7] and FFS [21] models trained with and without FEVER-OOD with PASCAL VOC as in-distribution are shown in Fig. 15 for OpenImages [22] as OOD, and in Fig. 16 fos MS-COCO [28] as OOD.

13. Limitations and Potential Negative Impact

This section discusses some limitations and potential negative impact of FEVER-OOD, identifying the following:

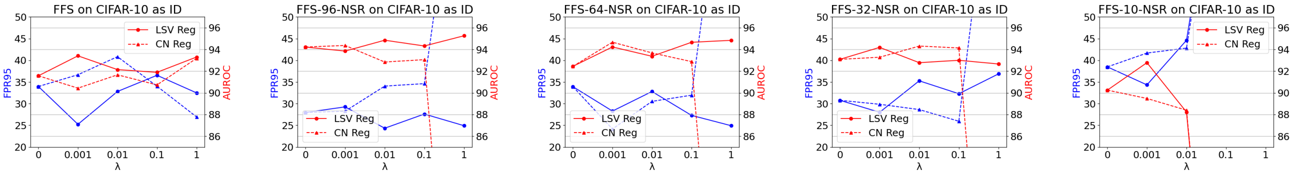
- FEVER-OOD does not entirely avoid the null space vulnerabilities. While we reduce the size of it, there might be some anomalies with large components in the feature space.
- Careful fine tuning is needed in some instances, specially when reducing the null space significantly. We did not identify any condition to estimate the regularizer weight *a priori*.
- While our analysis show a large change in Energy for far anomalies (OOD samples), we did not test the performance of FEVER-OOD in this cases.

Finally, our work might have some potential negative impact. For instance, the exploration of these vulnerabilities might allow for tailored generated outliers that fool (in an adversarial sense) models based on free energy OOD detection. Additionally, we trained several models with different in-distribution datasets to perform the ablation studies, having a negative environmental effect due to the large power consumption for GPU training for such extensive studies.



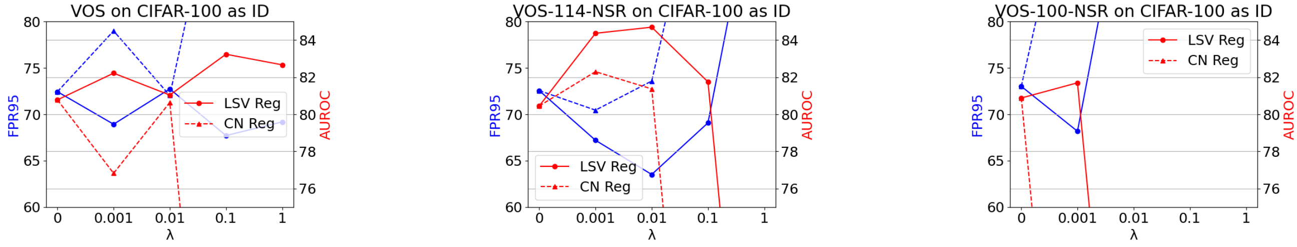
(a) VOS (b) VOS-96-NSR (c) VOS-64-NSR (d) VOS-32-NSR (e) VOS-10-NSR

Figure 10. Ablations of the loss weight for the LSVR and CNR in FEVER-OOD for VOS, using CIFAR-10 as in-distribution (ID).



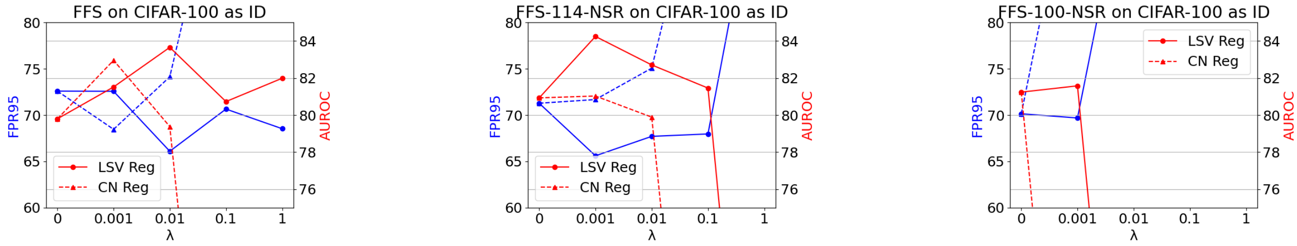
(a) FFS (b) FFS-96-NSR (c) FFS-64-NSR (d) FFS-32-NSR (e) FFS-10-NSR

Figure 11. Ablations of the loss weight for the LSVR and CNR in FEVER-OOD for FFS, using CIFAR-10 as in-distribution (ID).



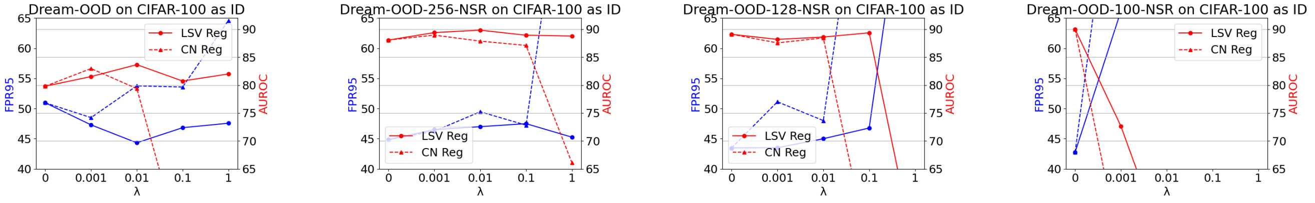
(a) VOS (b) VOS-114-NSR (c) VOS-100-NSR

Figure 12. Ablations of the loss weight for the LSVR and CNR in FEVER-OOD for VOS, using CIFAR-100 as in-distribution (ID).



(a) FFS (b) FFS-114-NSR (c) FFS-100-NSR

Figure 13. Ablations of the loss weight for the LSVR and CNR in FEVER-OOD for FFS, using CIFAR-100 as in-distribution (ID).



(a) Dream-OOD (b) Dream-OOD-256-NSR (c) Dream-OOD-128-NSR (d) Dream-OOD-100-NSR

Figure 14. Ablations of the loss weight for the LSVR and CNR in FEVER-OOD for Dream-OOD, using CIFAR-100 as in-distribution (ID).

Table 8. CIFAR-10 Results (ID Acc. = in-distribution accuracy; Non-energy based methods).

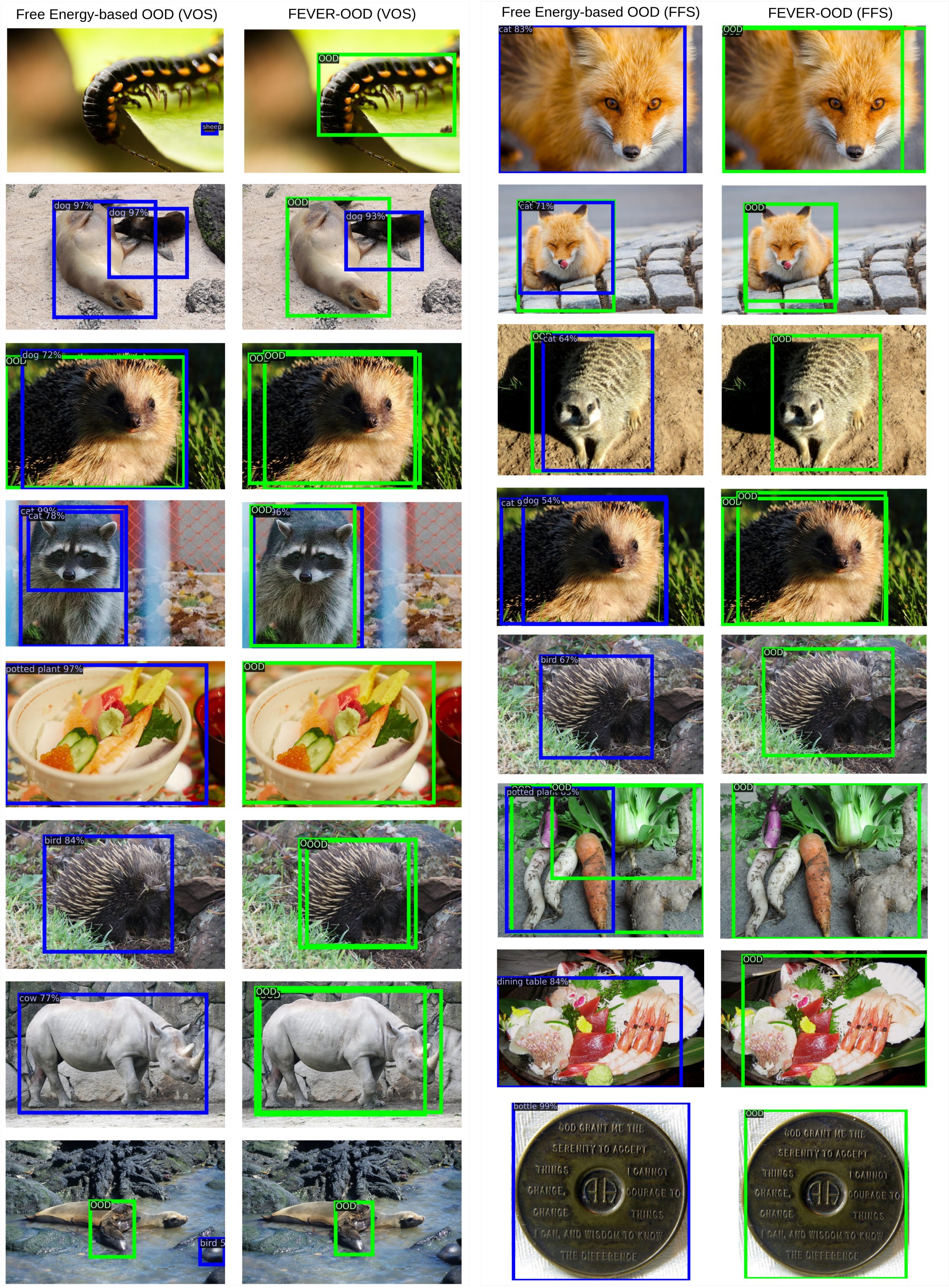
Non-energy based method	OOD Datasets												ID Acc
	Texture		SVHN		Place365		LSUN		iSUN		Avg		
	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	
ViM [41]	24.35	95.20	24.95	95.36	44.70	90.71	18.80	96.63	29.25	95.10	28.41	94.60	94.21
ODIN [27]	56.40	86.21	20.93	95.55	63.04	86.57	7.26	98.53	33.17	94.65	36.16	92.30	94.21
Softmax [15]	66.45	88.50	59.66	91.25	62.46	88.64	45.21	93.80	54.57	92.12	57.67	90.86	94.21
GradNorm[17]	71.66	80.79	80.86	81.41	80.71	72.57	53.87	88.39	60.32	88.00	69.49	82.23	94.21
KNN [38]	27.57	94.71	24.53	95.96	50.90	89.14	25.29	95.69	25.55	95.26	30.77	94.15	94.21
NPOS [39]	8.39	94.67	5.61	97.64	18.57	91.35	4.08	97.52	14.13	94.92	10.16	95.22	93.86

Table 9. CIFAR-100 Results (ID Acc. = in-distribution accuracy; Non-energy based methods).

Non-energy based method	OOD Datasets												ID Acc
	Texture		SVHN		Place365		LSUN		iSUN		Avg		
	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	
ViM [41]	86.00	71.95	54.30	88.85	84.70	74.64	57.15	88.17	56.65	87.13	67.76	82.15	73.12
ODIN [27]	85.75	73.17	89.50	76.13	41.50	91.60	74.70	83.93	90.20	68.27	76.33	78.62	73.12
Softmax [15]	86.45	71.32	85.30	72.41	73.40	81.09	85.55	74.00	88.55	68.59	83.85	73.48	73.12
GradNorm[17]	96.20	52.17	91.05	67.13	55.72	86.09	97.80	44.21	89.71	58.23	86.10	61.57	73.12
KNN [38]	88.00	67.19	66.38	83.76	79.17	71.91	70.96	83.71	77.83	78.85	76.47	77.08	73.12
NPOS [39]	33.07	92.86	17.98	96.43	80.41	73.74	28.90	92.99	43.50	89.56	40.77	89.12	73.78

Table 10. ImageNet-100 Results (ID Acc. = in-distribution accuracy; Non-energy based methods).

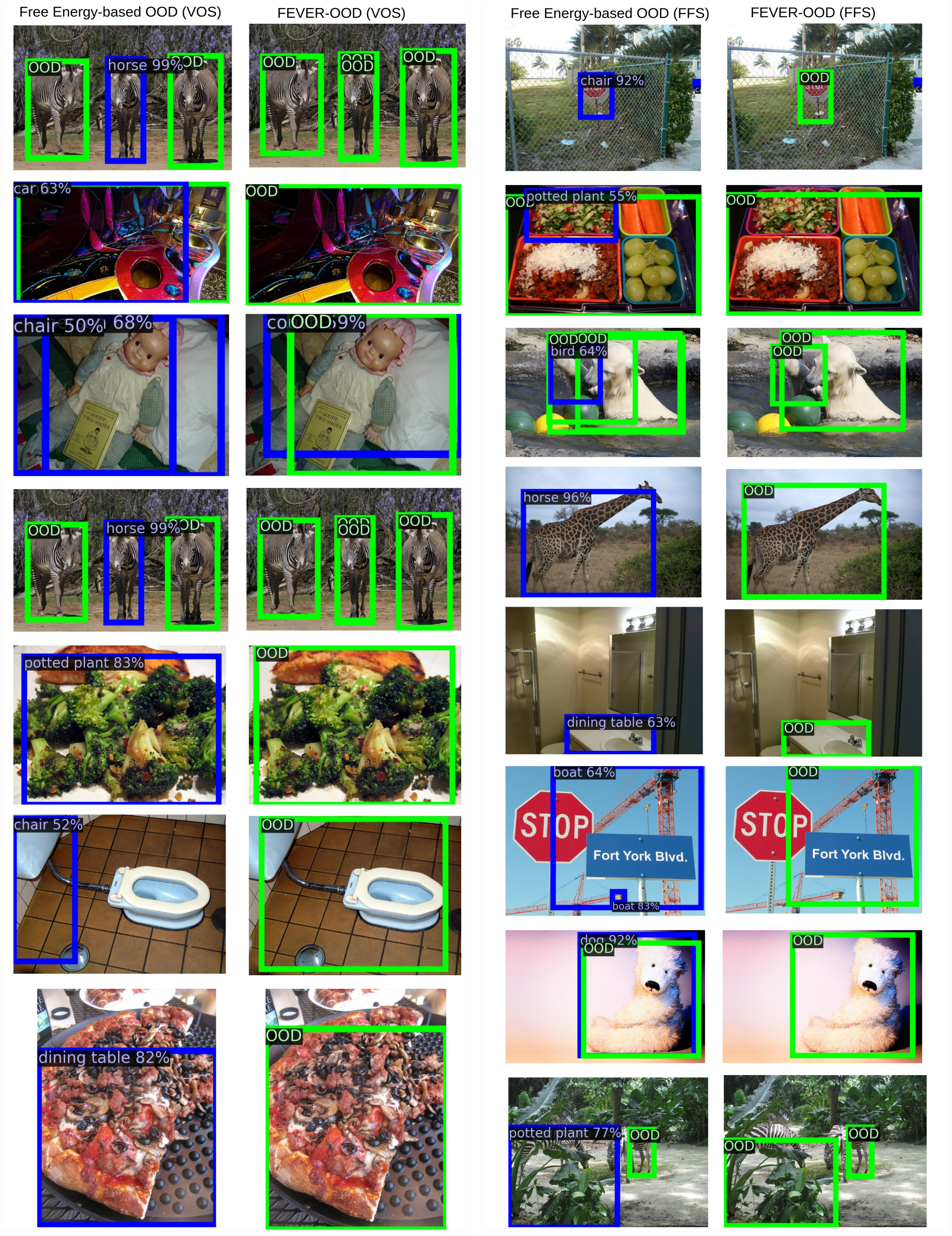
Non-energy based method	OOD Datasets										ID Acc
	iNaturalist		Place365		SUN		Textures		Avg		
	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	FPR95	AUROC	
ViM [41]	72.40	84.88	76.20	81.54	73.80	83.99	22.20	95.63	61.15	86.51	84.16
ODIN [27]	53.00	89.52	70.40	82.77	66.90	85.01	48.40	89.19	59.67	86.62	84.16
Softmax [15]	76.30	82.20	81.90	77.54	82.70	78.35	75.30	80.01	79.05	79.52	84.16
GradNorm[17]	50.82	84.86	68.27	74.46	65.77	77.11	40.48	88.17	56.33	81.15	84.16
KNN [38]	56.96	86.98	64.54	83.68	63.04	85.37	15.83	96.24	50.09	88.07	84.16
NPOS [39]	53.84	86.52	59.66	83.50	53.54	87.99	8.98	98.13	44.00	89.04	85.37



(a)

(b)

Figure 15. Additional visualization of detected objects on the OOD images (from OpenImages [22]) by free energy-based OOD (VOS) [7], free energy-based OOD (FFS) [21] and FEVER-OOD (our approach). The in-distribution is PASCAL VOC [10] dataset. Blue: OOD objects detected and mis-classified as being in-distribution. Green: the same OOD objects correctly detected as OOD by FEVER-OOD (ours).



(a) (b)

Figure 16. Additional visualization of detected objects on the OOD images (from MS-COCO [28]) by free energy-based OOD (VOS) [7], free energy-based OOD (FFS) [21] and FEVER-OOD (our approach). The in-distribution is PASCAL VOC [10] dataset. Blue: OOD objects detected and mis-classified as being in-distribution. Green: the same OOD objects correctly detected as OOD by FEVER-OOD (ours).