Learning Deep Classifiers Consistent with Fine-Grained Novelty Detection

NRI: FND: Towards Scalable and Self-Aware Robotic Perception

Nuno Vasconcelos (PI, UC San Diego), Jiacheng Cheng (PhD student, UC San Diego)

https://ieeexplore.ieee.org/document/9578373

Background:

Deep neural networks (DNNs) are widely deployed in different autonomous robotics systems.

- The *closed-world assumption* underlying the training of CNNs can be easily violated
- DNNs tend to assign examples from *novel/unseen* classes to one of its *training/seen* classes

Such mistaken predictions can lead to wrong decision made by robots.

Hence, the ability to perform *novelty detection* (ND) is indispensable for robust intelligent systems.

Challenges:

- DNN provides unreliable estimates for class-posterior probability

$$P_{Y|X}(y|v(x)) = \frac{\exp(\langle w_y, v(x)\rangle + b_y)}{\sum_k \exp(\langle w_k, v(x)\rangle + b_y)}$$

- (Probabilistic ND) Class-conditional density $P_{X|Y}(v(x)|y)$ is unidentifiable

$$P_{X|Y}(v(x)|y) = q(x)\mathrm{e}^{\langle w_y, v(x)\rangle - \psi(w_y)},$$ an exponential family distribution

- (Distance-based ND) Metric defining DNN embedding's geometry is unidentifiable

$$P_{X|Y}(v(x)|y) \propto_x \mathrm{e}^{-d_\phi(v(x),\,\mu_y)},$$ d_ϕ: the corresponding Bregman divergence

Solution:

Force the class-conditional distributions to be Gaussians by regularizing DNN training, i.e.,

$$P_{Y|X}(y|v(x)) = \mathrm{G}(\mathrm{v}(x); \mu_y, \Sigma) = \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}}\, e^{-d_\phi(v(x),\mu_y)},$$

Where $d_\phi\big(v(x), \mu_y\big) = ||\mathrm{v}(x) - \mu_y||^2_\Sigma$ is the Mahalanobis distance.

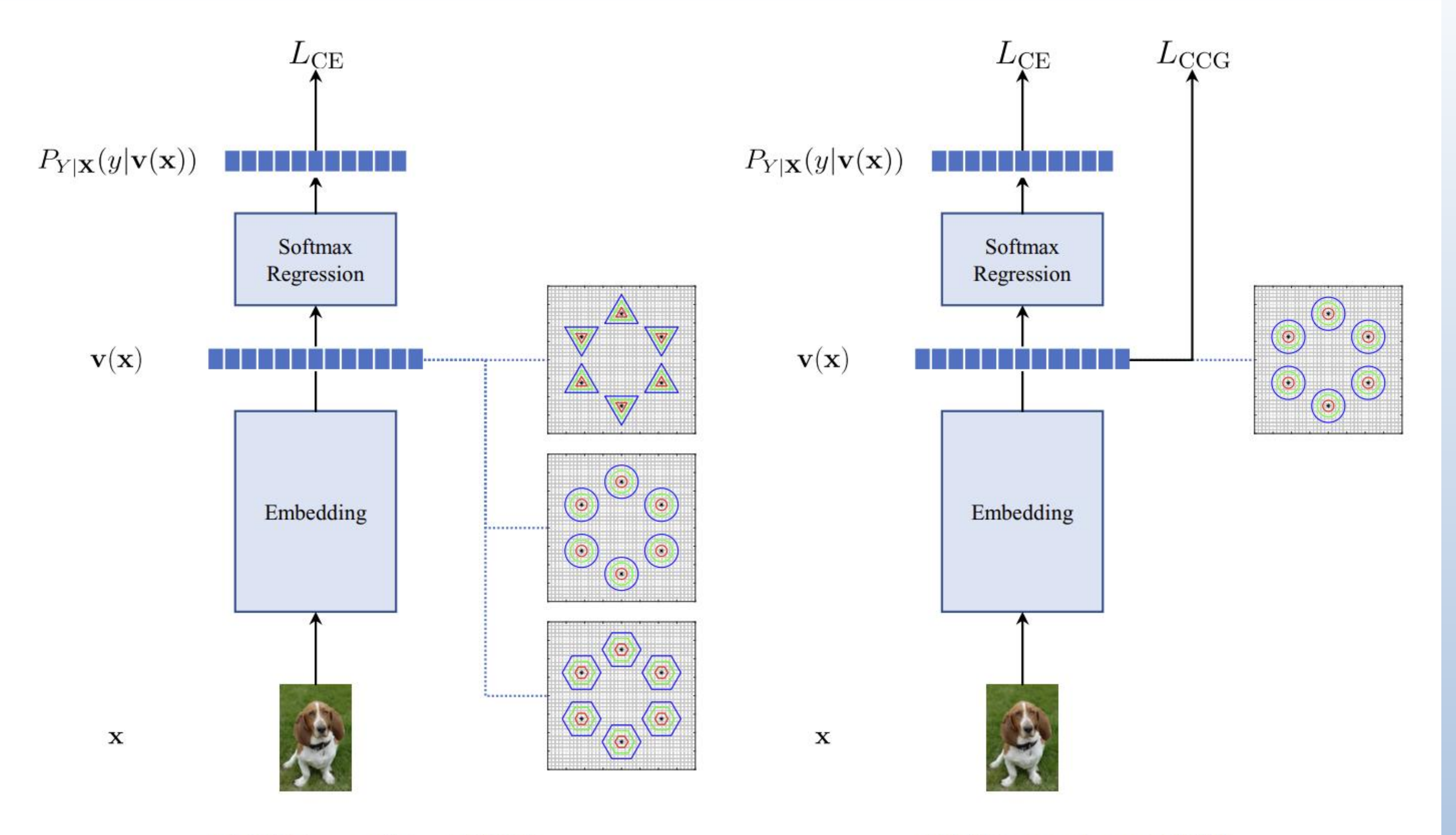
A Class-conditional Gaussianity (CCG) loss was proposed to supplement the standard cross-entropy loss.

$$L_{CE} + \lambda\, L_{CCG}$$

Novelty detection is performed by thresholding the Mahalanobis distance

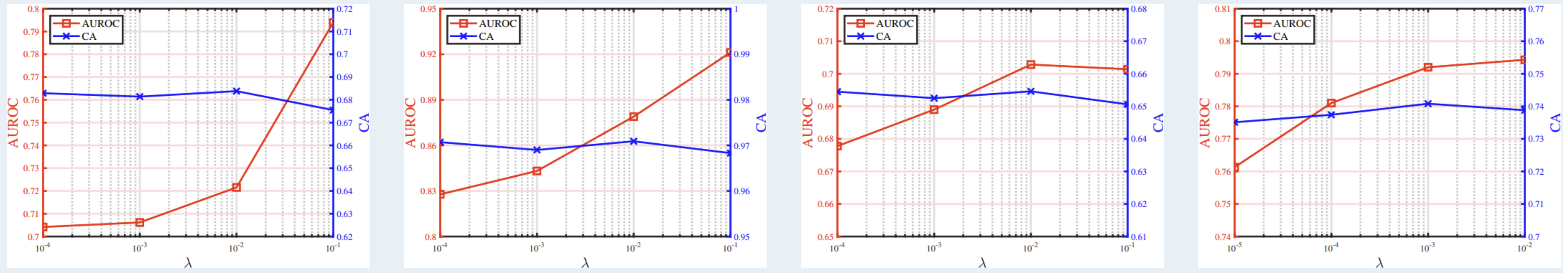
$$Novelty(x) = \min_y ||\mathrm{v}(x) - \mu_y||^2_\Sigma$$

Reference: J. Cheng, N. Vasconcelos, "Learning deep classifiers consistent with fine-grained novelty detection", IEEE/CVF CVPR 2021, DOI: 10.1109/CVPR46437.2021.00171.



(a) ND-inconsistent CNN (b) ND-consistent CNN

(a): CNN trained with the cross-entropy loss L_{CE} is inconsistent with ND. Because the class-conditional distributions learned by the CNN are unidentifiable, multiple sets of distributions (visualized using contour plots) are compatible with the CNN parameters. (b): Regularization with the proposed CCG loss L_{CCG} makes the distributions identifiable, in fact Gaussians.



(a) Stanford Dogs (b) FounderType-200 (c) CUB-200-2010 (d) Caltech-256

Figure 3: AUROC and closed-world classification accuracy (CA) versus λ.

Scientific Impact:

- A theoretical analysis of the softmax classifier, showing that although it learns exponential family class-conditional distributions, these are not identifiable.
- the derivation of identifiability conditions, that guarantee Gaussian distributions.
- the CCG regularization loss that encourages these conditions to hold, producing classifiers that are consistent with novelty detection.
- evaluations on various fine-grained visual classification datasets demonstrate that our proposed method significantly advances the state-of-the-art for novelty detection