Self-Feature Distillation with Uncertainty Modeling for Degraded Image Recognition

Zhou Yang[1] Weisheng Dong[1*] Xin Li[2] Jinjian Wu[1] LeiDa Li[1] Guangming Shi[1]

[1]School of Artificial Intelligence, Xidian University [2]Lane Dep. of CSEE, West Virginia University

ECCV TEL AVIV 2022

Motivation

- Shrinking the distribution distance between degraded and high-quality features is an effective way to improve the robustness of image recognition models.
- The commonly used feature reconstruction loss MSE (i.e., the L2-norm) potentially treats the variance of each position in the feature map as a constant, which is not suitable due to the diversity of degradation.

Contributions

- **A new State-of-the-art method for degraded images recognition.**
- We modeled the uncertainty of recognition problems under various types of image degradation.
- We proposed a self-features distillation learning schedule, a weighted and regularized L2-norm loss function $\frac{\|z-\hat{z}\|^2}{\theta^2}+log\theta^2$ for distillation. The mean (high quality like features) and variance (uncertainty) of the proposed method were learned by DCNNs.

Uncertainty model for LQ images recognition

- Given a degraded image $\tilde{x}$, the recognition problem can be formulated as a MAP estimation problem, i.e.,

$$\operatorname{argmax} p(z, y \mid \tilde{x}) = \operatorname{argmax} p(y \mid z, \tilde{x}) p(z \mid \tilde{x})$$

where z is the high-quality feature and y is the label.

- we can formulate the observation model with the estimated HQ-like feature $\hat{z}$ and the target HQ feature z as a **Gaussian likelihood function**: $z = \hat{z} + \varepsilon \cdot \theta$, $p(z|\tilde{x}) = p(z, \theta|\tilde{x})$, so we have:

$$\log p(z_i, \theta_i \mid \tilde{x}_i) = -\frac{\|z_i - g(\tilde{x}_i; \Theta_2)\|^2}{2\theta_i^2} - \log \theta_i^2 \longrightarrow L_{ULFD}$$

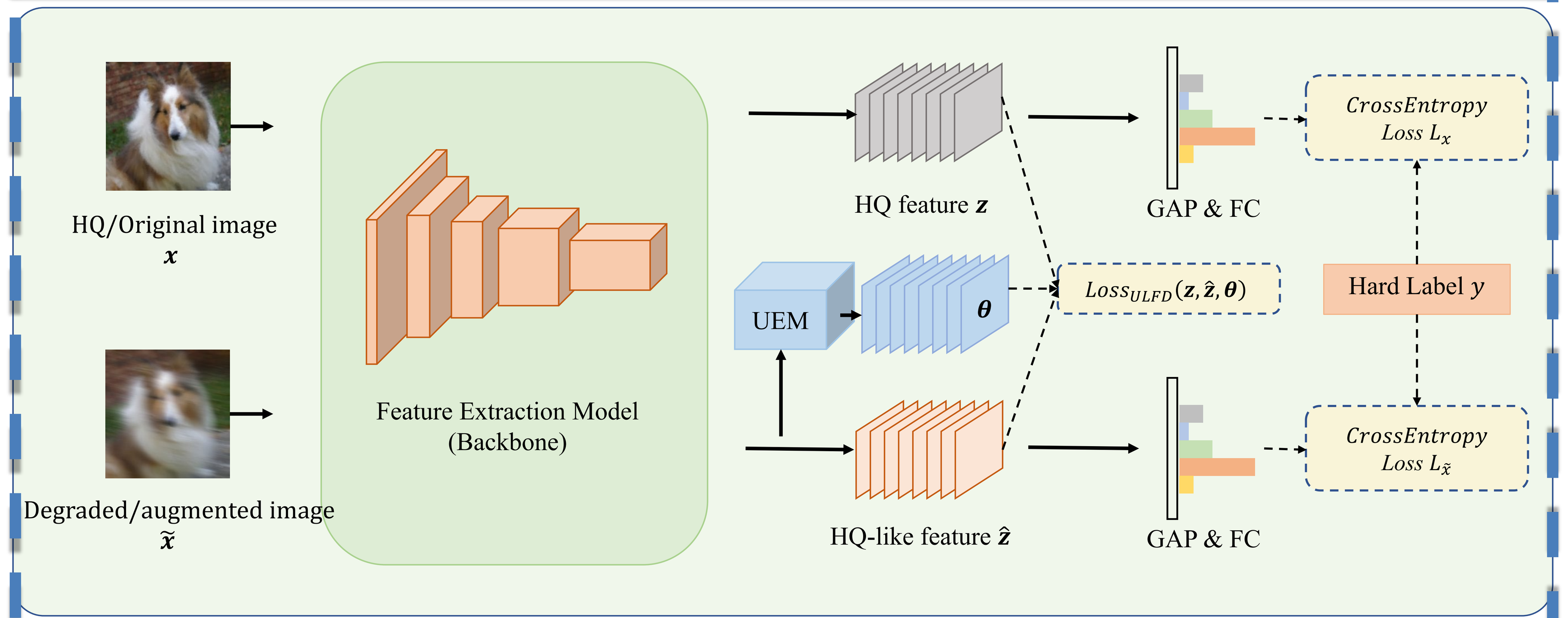
$g(\cdot)$ is the backbone and Θ_2 is the parameters.

- Then the tototal loss can be formulated as :

$$L = L_{CE} + \lambda \cdot L_{ULFD}$$

- To implement the above idea, we add a new branch (UEM) at the end of the backbone network to estimate the uncertainty θ.

Architecture of the proposed method



Simulation results

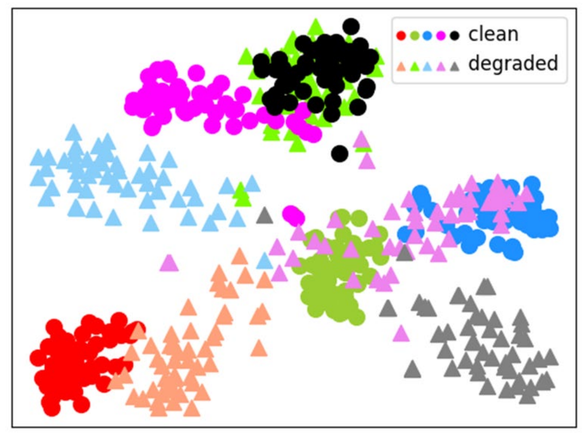
Methods	Architecture	HQ ↑	Seen ↑	Unseen ↑	mCE ↓
Vanilla [16]	ResNet50	**76.82%**	39.17%	47.11%	76.5%
DDP [46]		72.15%	48.21%	50.73%	62.78%
URIE [42]		73.80%	55.10%	56.50%	55.70%
KD VID [1]		74.85%	-	-	51.29%
QualNet50 [25]		75.43%	61.08%	58.10%	50.34%
Ours w/o UEM		75.81%	61.65%	60.23%	49.50%
Ours		76.23%	**63.44%**	**62.90%**	**46.37%**
Vanilla [47]	ResNeXt101	**79.68%**	47.08%	55.53%	69.76%
QualNet101 [25]		77.81%	65.47%	63.28%	42.61%
Ours w/o UEM		78.35%	66.81%	65.30%	41.23%
Ours		79.04%	**69.16%**	**67.83%**	**39.50%**

- The top-1 accuracy on HQ ImageNet-1K validation set, 15 types seen corrupted and 4 types unseen corrupted images in ImageNet-C validation set .

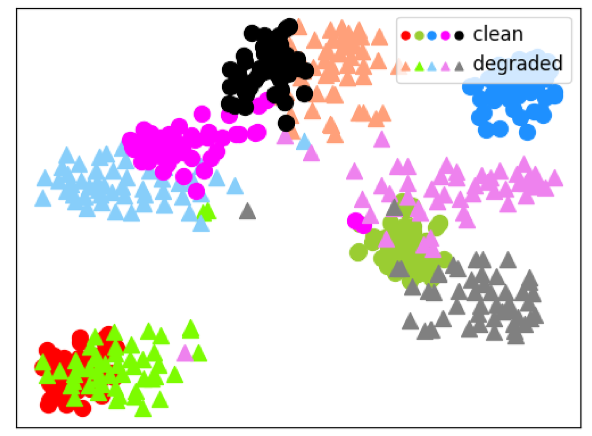
Methods	Architecture	Top-1 Accuracy ↑ Speckle-Noise	Gaussian-Blur	Spatter	Saturate
Vanilla [16]	ResNet50	35.49%	49.16%	41.87%	61.92%
QualNet50 [25]		63.50%	52.59%	54.56%	61.75%
Ours w/o UEM		65.25%	55.39%	56.33%	63.95%
Ours		**66.44%**	**58.59%**	**58.65%**	**67.92%**
Vanilla [47]	ResNeXt101	47.92%	57.94%	48.72%	67.52%
QualNeXt101 [25]		64.21%	57.24%	62.48%	69.19%
Ours w/o UEM		68.70%	61.25%	60.37%	70.86%
Ours		**71.23%**	**64.87%**	**63.04%**	**72.18%**

- Top-1 accuracy on 4 unseen corruptions in ImageNet-C validation set for robustness test.

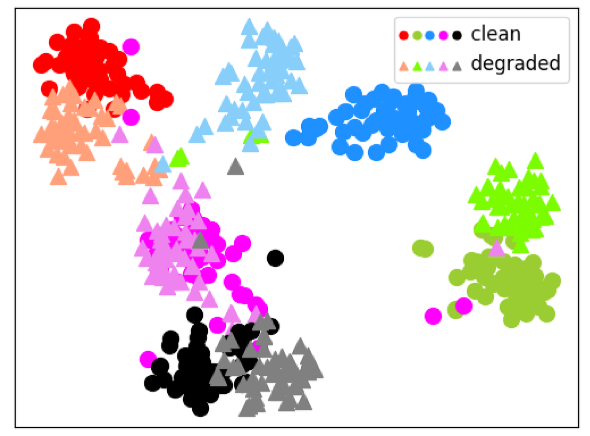
The visualization of t-SNE features



(a) Teacher



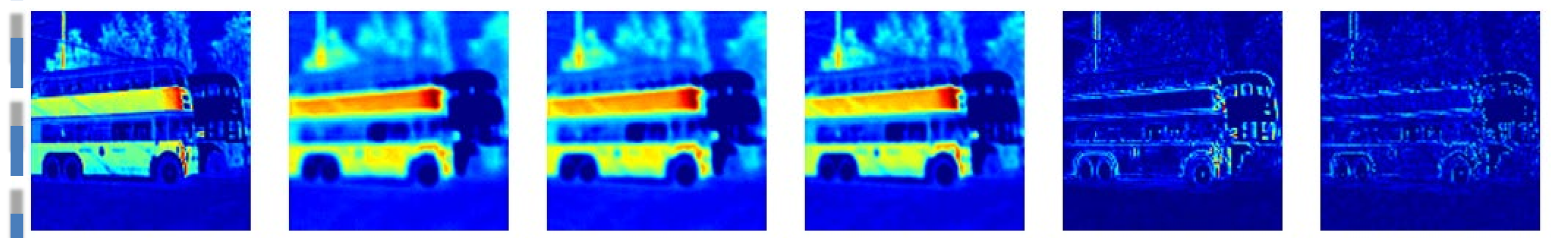
(b) QualNet



(c) Ours

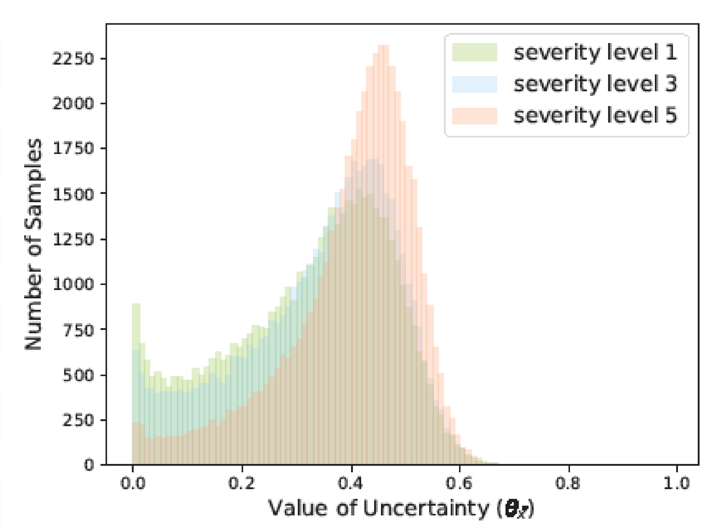
- The colored symbols represent the feature vectors extracted from the corresponding images. Symbols with the same color are from the same class.

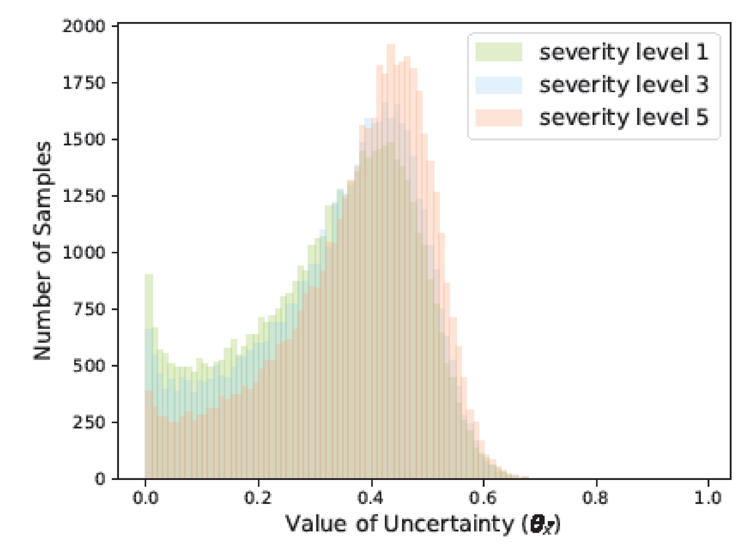
The visualization of reconstructed features



(a) HQ (b) LQ (c) DDP (d) Ours (e) Diff.DDP (f) Diff.Ours

Analysis of uncertainty learning





Code: https://github.com/yangzhou321/Distillation_with_UEM