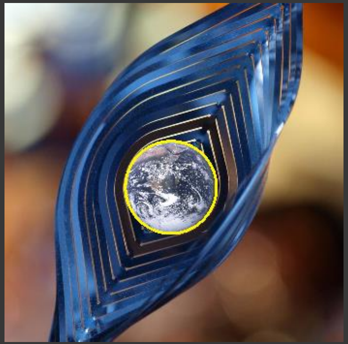
OpenGAN: Open Set Recognition via Open Data Generation



Shu Kong Deva Ramanan

Carnegie Mellon University

ARGO AI

Motivation

Autonomous vehicles invariably encounters *unknown* objects in the real open world, and it's crucial to detect them.

Machine-learned models

- Trained on a *closed-set*, e.g., training a K-way classifier on a dataset that has K classes of data.
- Tested in the real open world, which contains *unknown* examples outside the K classes

[Tesla crashes directly into overturned truck]

Problem: Open-Set Recognition

Open-Set Recognition: detecting the unknown through the lens of image classification.

- Learning a K-way classifier for data from the closed-set K classes
- Testing it on examples that contain *open-set* data that is from some unknown-classes.
  - ref. anomaly detection, out-of-distribution detection, etc.

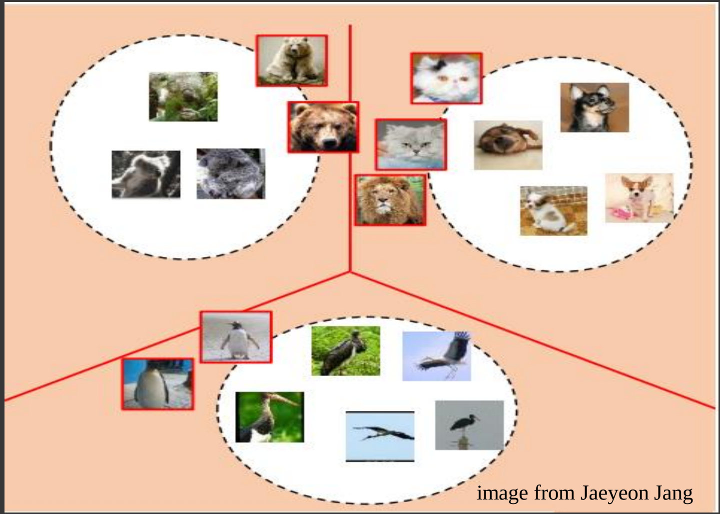
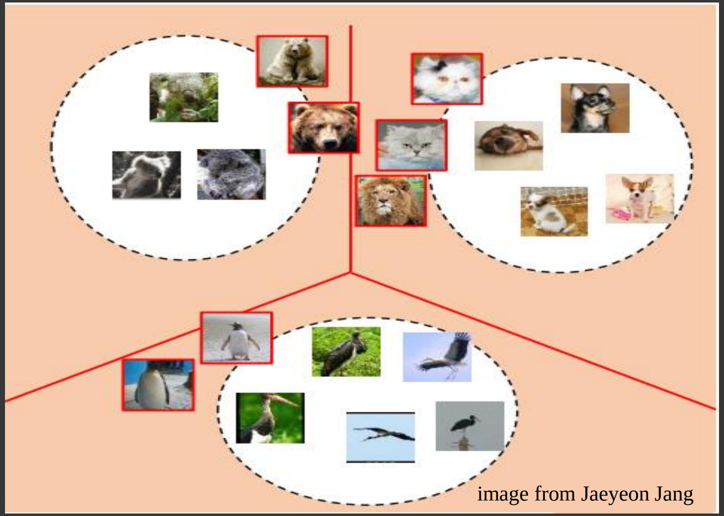


image from Jaeyeon Jang

Scheirer and Boult, “Toward Open Set Recognition”, PAMI 2012

Contributions

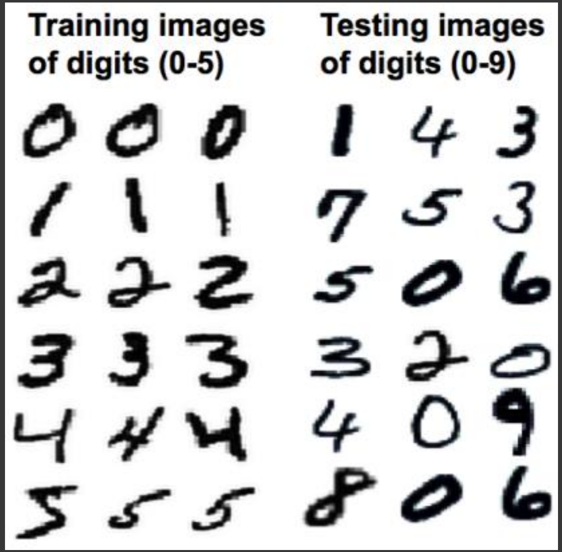
- *Method* OpenGAN, a lightweight (2MB) model atop the K-way classification network to recognize the open-set.
- *Performance* state-of-the-art performance under different setups, significantly better than prior methods.
- *Protocol* a realistic protocol for open-set recognition through the lens of semantic segmentation.



Status Quo

- *Typical* setup: splitting MNIST digits
  - closed-set: 0-5
  - open-set: 6-9

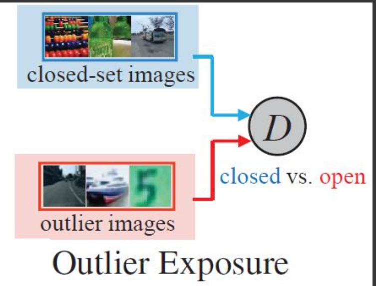
- Methods: uncertainty for open-set likelihood, as simple as *Max-Softmax Probability* (*MSP*)



Hendrycks and Gimpel, “A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks”, ICLR 2017

What if we embrace outliers in training?

- Outlier Exposure: use outlier data to train a closed-vs-open classifier [Hendrycks et al. ICLR 2019]
  Although it's an *atypical* setup, it's realistic because there are outliers / out-of-domain data in the wild.
  Yet, it sometimes performs poorly as the outliers do not span the whole open world.

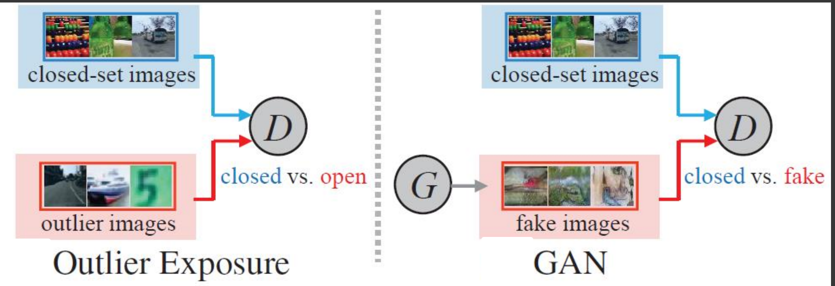


Dhamija et al., "Reducing Network Agnostophobia", NeurIPS, 2018
Hendrycks et al., "Deep Anomaly Detection with Outlier Exposure", ICLR, 2019

Generate fake open-set examples in training

- Outlier Exposure: use outlier data to train a closed-vs-open classifier [Hendrycks et al. ICLR 2019]
  Although it's an *atypical* setup, it's realistic because there are outliers / out-of-domain data in the wild.
  Yet, it sometimes performs poorly as the outliers do not span the whole open world.

- Without sufficient outliers, we synthesize outliers as fake open-set data? Using a GAN?
  Using GAN-discriminator as the open-set likelihood, as it learns closed-set data distribution!
  But it does not work due to instable training of GANs. Recall GAN-discriminator ideally should be confused by closed-set and fake open-set



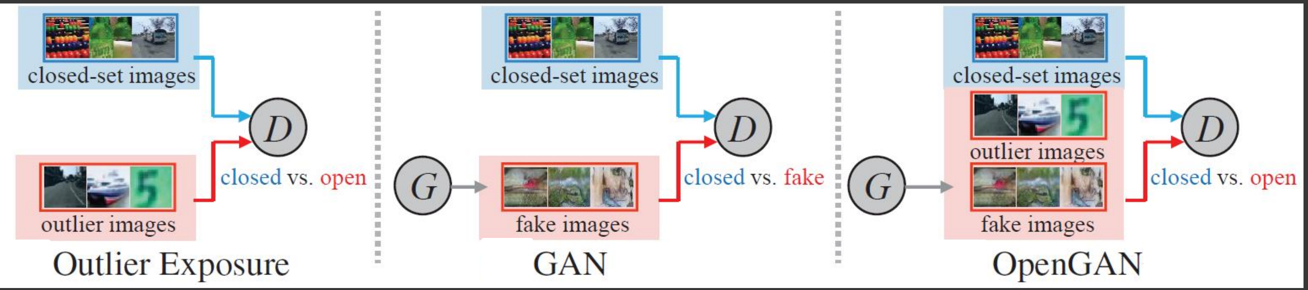
Schlegl et al. "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery", IPMI, 2017
Neal et al. "Open set learning with counterfactual images", ECCV, 2018
Zenati et al., "Adversarially learned anomaly detection", ICDM, 2018

OpenGAN ≈ OutlierExposure + GAN

- Outlier Exposure: use outlier data to train a closed-vs-open classifier [Hendrycks et al. ICLR 2019]
  Although it's an *atypical* setup, it's realistic because there are outliers / out-of-domain data in the wild.
  Yet, it sometimes performs poorly as the outliers do not span the whole open world.

- Without sufficient outliers, we synthesize outliers as fake open-set data? Using a GAN?
  Using GAN-discriminator as the open-set likelihood, as it learns closed-set data distribution!
  But it does not work due to instable training of GANs. Recall GAN-discriminator ideally should be confused by closed-set and fake open-set

- OpenGAN augments training outliers with synthesized data
  - ~ GAN, it repurposes GAN-discriminator as the open-set likelihood function *D*, synthesizes data to better span the open world.
  - ~ Outlier Exposure, it uses outliers to train an open-set classifier *D*, stabilize training of GANs, select right closed-vs-open classifier *D*

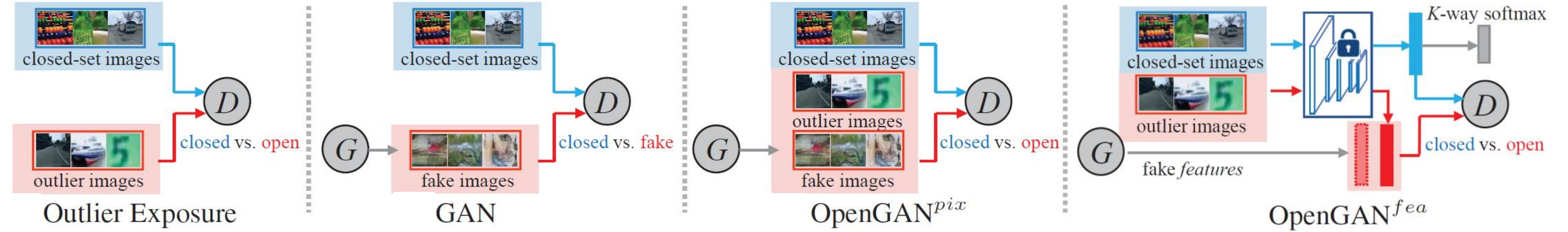


OpenGAN^{fea}

- *features > pixels*.
  - It's more efficient to generate features than pixel images.
  - This enables closed-world systems to be readily modified for open-set recognition.

- *discriminator > generator*.
  - Recall GANs mainly focus on the generator and generating images.
  - Our goal is to learn a robust open-vs-closed *discriminator* rather than generating realistic pixel images.

- *classification > reconstruction*.
  - Existing methods commonly use reconstruction errors for open-set detection, but it's challenging to reconstruct high-res images.
  - We directly use the GAN-discriminator as the open-set likelihood, allowing open-set recognition over high-res images



Experiments

- Three different settings (detailed later)

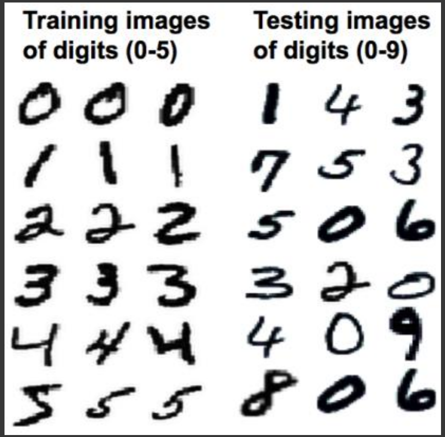
- Metrics
  - Open-set *detection*: area under ROC curve (AUROC)
  - Open-set *recognition*: macro-average F1 score over (K+1) classes

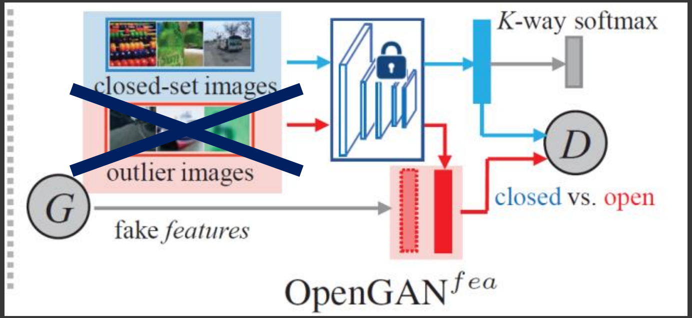
- Compared methods

Baselines	Likelihoods	Bayesian Networks	GANs	State-of-the-art
• Nearest Neighbors [SIGMOD2000] • Gaussian Mixture Model [arxiv2021] • CLS open-vs-closed classifier (ref. Outlier Exposure) [ICLR2017] • (K+1)-way classifier [PASCAL, IJCV2015]	• Max Softmax Prob. (MSP) [ICLR2017] • Entropy [NeurIPS2016] • calibrated MSP (MSPc) [NeurIPS2016] • OpenMax [CVPR2016] • C2AE [CVPR2019] • Gaussian Discrim. Model (GDM) [NeurIPS2018]	• Monte Carlo est. (MCdrop) [ICML2016]	• G-OpenMax [BMVC2017] • OSRCI [ECCV2018] • BiGAN [ICDM2018]	• PRL [ECCV2020] • Hybrid [ECCV2020] • GDFR [CVPR2020] • CROSR [CVPR2019]

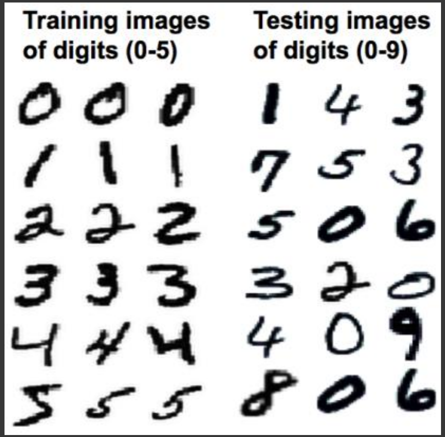
Setup-I: Single Dataset Split

- Protocol
  - split a single dataset into the closed- and open-sets w.r.t class labels
  - train on the closed-train-set only
  - Validate on closed- and open-sets
  - measure open-set detection performance using AUROC.
- Datasets
  - CIFAR / MNIST / SVHN: 6 random classes as the closed-set, the remaining 4 as the open-set
  - TinyImageNet: 20 random classes as the closed-set, the remaining 180 as the open-set
- Note
  - We do not use outliers to train OpenGAN, hence we have a typical GAN-discriminator, denoted as OpenGAN-0





Setup-I: Single Dataset Split

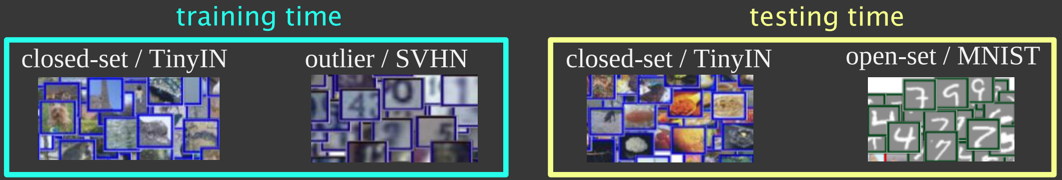


- Protocol
  - split a single dataset into the closed- and open-sets w.r.t class labels
  - train on the closed-train-set only
  - Validate on closed- and open-sets
  - measure open-set detection performance using AUROC.
- Datasets
  - CIFAR / MNIST / SVHN: 6 random classes as the closed-set, the remaining 4 as the open-set
  - TinyImageNet: 20 random classes as the closed-set, the remaining 180 as the open-set
- Note
  - We do not use outliers to train OpenGAN, hence we have a typical GAN-discriminator, denoted as OpenGAN-0
- Salient conclusions
  - OpenGAN-0 clearly performs the best
  - Methods (e.g., NN and OpenGAN) perform much better on off-the-shelf features than raw pixels

Dataset	MSP [24]	MSP_c [29]	MCdrop [17]	GDM [28]	OpenMax [5]	GOpenMax [18]*	OSRCI [33]*	C2AE [35]*	CROSR [54]*	RPL [10]*	Hybrid [57]*	GDFR [37]*	NN^{pix} [41]	NN^{fea} [41]	OpenGAN -0^{pix}	OpenGAN -0^{fea}
MNIST	.977	.985	.984	.989	.981	.984	.988	.989	.991	.996	.995	—	.931	.981	.987	**.999**
SVHN	.886	.891	.884	.866	.894	.896	.910	.922	.899	.968	.947	.935	.534	.888	.881	**.988**
CIFAR	.757	.808	.732	.752	.811	.675	.699	.895	.883	.901	.950	.807	.544	.801	.971	**.973**
TinyImgNet	.577	.713	.675	.712	.576	.580	.586	.748	.589	.809	.793	.608	.528	.692	.795	**.907**

Setup-II: Cross-Dataset Evaluation

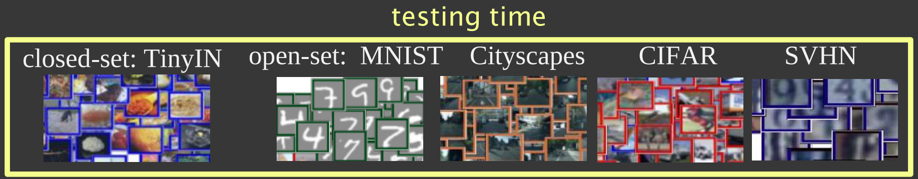
- Protocol: a less biased protocol that uses cross-dataset images as the train, val, and test sets.
- Datasets: TinyImageNet as the 200-class closed-set, open-set / outliers from {MNIST, CIFAR, SVHN, Cityscapes}



Shafaei et al., "A Less Biased Evaluation of Out-of-distribution Sample Detectors", BMVC, 2018

Setup-II: Cross-Dataset Evaluation

- Protocol: a less biased protocol that uses cross-dataset images as the train, val, and test sets.
- Datasets: TinyImageNet as the 200-class closed-set, open-set / outliers from {MNIST, CIFAR, SVHN, Cityscapes}

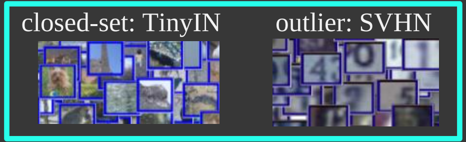


Shafaei et al., "A Less Biased Evaluation of Out-of-distribution Sample Detectors", BMVC, 2018

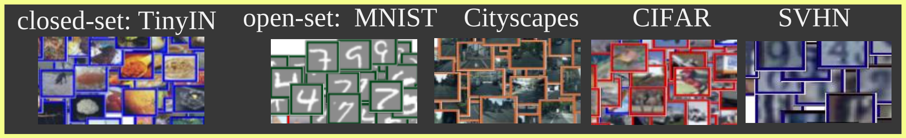
Setup-II: Cross-Dataset Evaluation

- Protocol: a less biased protocol that uses cross-dataset images as the train, val, and test sets.
- Datasets: TinyImageNet as the 200-class closed-set, open-set / outliers from {MNIST, CIFAR, SVHN, Cityscapes}

training time



testing time

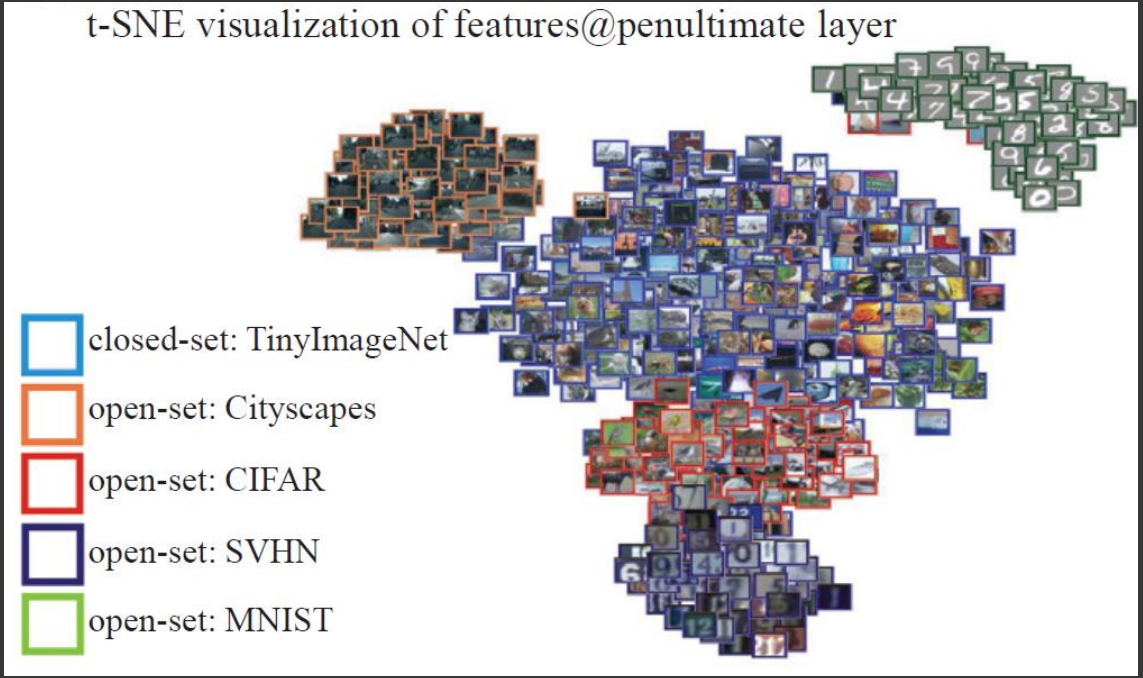


- Salient conclusions
  - Methods perform much better on off-the-shelf features rather than pixels, ref OpenGAN and CLS.
  - CLS ^{fea} already outperforms prior methods when trained on features, though CLS ^{pix} performs poorly.
  - (K+1)-way model works quite well, but OpenGAN performs the best.

open-test	*metric*	MSP [24]	OpenMax [5]	NN^{fea} [41]	GMM [26]	C2AE [35]	MSP$_c$ [29]	MCdrop [17]	GDM [28]	CLS^{pix}	(K+1)	CLS^{fea}	Open GAN^{pix}	Open GAN^{fea}
CIFAR	AUROC	.769$^{.000}$	.669$^{.011}$	.927$^{.000}$	.961$^{.013}$	.767$^{.020}$	.791$^{.007}$	.809$^{.005}$	.961$^{.007}$	.754$^{.367}$	.880$^{.091}$	.928$^{.113}$	.981$^{.027}$	.980$^{.011}$
	F1	.548$^{.002}$	.507$^{.001}$	.525$^{.000}$	.544$^{.002}$	.564$^{.002}$	.553$^{.003}$	.564$^{.001}$	.519$^{.003}$	.545$^{.032}$	.558$^{.017}$	.555$^{.027}$	.563$^{.035}$	.585$^{.003}$
SVHN	AUROC	.695$^{.000}$	.691$^{.014}$	.994$^{.000}$	.990$^{.016}$	.657$^{.018}$	.863$^{.013}$	.783$^{.009}$	.999$^{.006}$	.701$^{.224}$	.948$^{.068}$	.955$^{.052}$	.980$^{.014}$	.991$^{.013}$
	F1	.567$^{.002}$	.551$^{.002}$	.545$^{.000}$	.574$^{.002}$	.565$^{.001}$	.572$^{.002}$	.572$^{.001}$	.575$^{.002}$	.572$^{.027}$	.564$^{.015}$	.578$^{.014}$	.574$^{.009}$	.583$^{.008}$
MNIST	AUROC	.764$^{.000}$	.690$^{.019}$	.901$^{.000}$	.964$^{.021}$	.755$^{.008}$	.832$^{.017}$	.801$^{.009}$	.957$^{.007}$	.986$^{.327}$	.944$^{.015}$	.961$^{.083}$	.983$^{.068}$	.989$^{.014}$
	F1	.559$^{.001}$	.536$^{.013}$	.553$^{.000}$	.547$^{.008}$	.575$^{.001}$	.564$^{.001}$	.563$^{.001}$	.552$^{.002}$	.565$^{.020}$	.586$^{.021}$	.583$^{.010}$	.569$^{.016}$	.582$^{.005}$
Citysc.	AUROC	.789$^{.000}$	.693$^{.021}$	.715$^{.000}$	.867$^{.016}$	.814$^{.010}$	.851$^{.003}$	.868$^{.003}$	.513$^{.005}$	.646$^{.332}$	.971$^{.050}$	.828$^{.032}$	.933$^{.026}$	.978$^{.013}$
	F1	.579$^{.002}$	.514$^{.002}$	.583$^{.000}$	.572$^{.003}$	.589$^{.002}$	.583$^{.001}$	.571$^{.001}$	.546$^{.003}$	.589$^{.007}$	.561$^{.029}$	.587$^{.006}$	.588$^{.007}$	.587$^{.000}$
average	AUROC	.754	.686	.884	.945	.748	.834	.815	.857	.772	.936	.918	.969	**.984**
	F1	.560	.527	.552	.559	.569	.568	.567	.548	.568	.565	.576	.573	**.584**

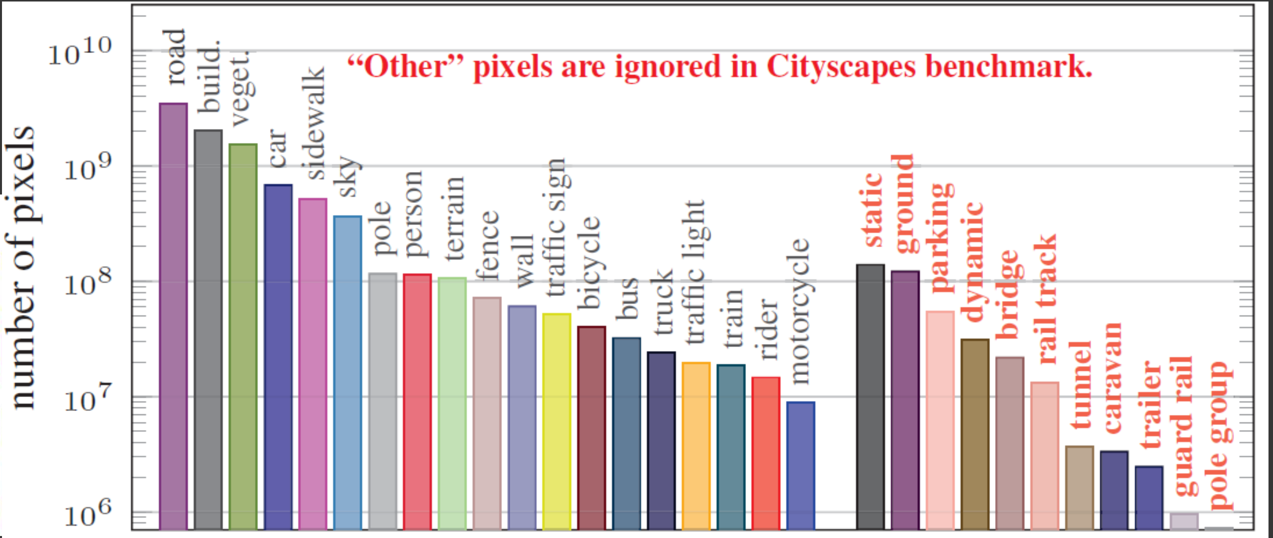
Setup-II: Cross-Dataset Evaluation

Visually explaining why OpenGAN ^{fea} works



Setup-III: Open-Set Semantic Segmentation

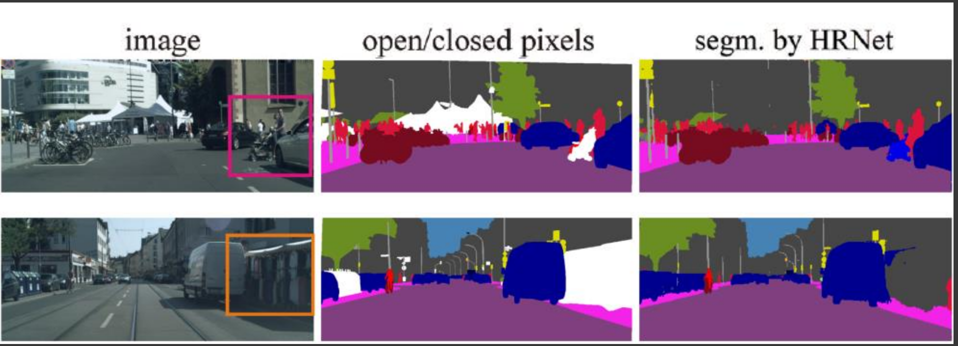
- Cityscapes, as well as other semantic segmentation datasets, ignores many “other” pixels in benchmarking.
- Importantly, ignored pixels can be from vulnerable objects such as strollers / wheelchairs.
- Semantic segmentation networks did not train on these ignored pixels, which then become the open-set.
- A realistic protocol for open-set recognition:
  - repurposing ignored pixels as training outliers and the testing open-set.
  - splitting Cityscapes trainset into our-trainset (2,965 images) and our-valset (10 images),
  - using Cityscapes valset as our-testset (500 images).



Setup-III: Open-Set Semantic Segmentation

- Cityscapes, as well as other semantic segmentation datasets, ignores many “other” pixels in benchmarking.
- Importantly, ignored pixels can be from vulnerable objects such as strollers / wheelchairs.
- Semantic segmentation networks did not train on these ignored pixels, which then become the open-set.
- A realistic protocol for open-set recognition:
  - repurposing ignored pixels as training outliers and the testing open-set.
  - splitting Cityscapes trainset into our-trainset (2,965 images) and our-valset (10 images),
  - using Cityscapes valset as our-testset (500 images).

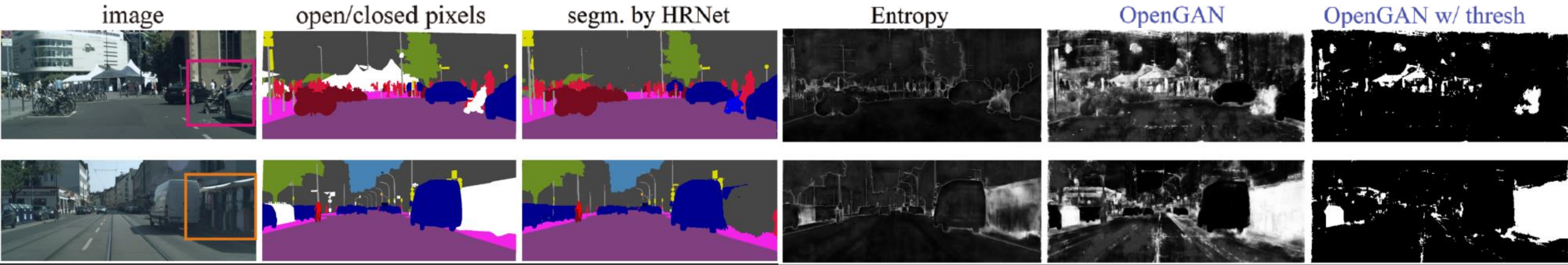
[Wang et al. TPAMI 2019]



Setup-III: Open-Set Semantic Segmentation

- $(K+1)$-way HRNet works quite well. [Wang et al. TPAMI 2019]
- OpenGAN^{fea} performs the best, presumably owing to more balanced sampling between the closed-set pixels and outliers in training
- The street-market is *a real open-set example* – there is not another similar street-market that is as big or selling clothes.

MSP [24]	Entropy [49]	OpenMax [5]	C2AE [35]	MSP$_c$ [29]	MCdrop [17]	GDM [28]	GMM [26]	HRNet-$(K+1)$	OpenGAN-0^{fea}	CLS^{fea}	OpenGAN^{fea}
.721	.697	.751	.722	.755	.767	.743	.765	.755	.709	.861	**.885**



Setup-III: Open-Set Semantic Segmentation

Visualization of the synthesized images by OpenGAN

- For OpenGAN ^{pix}, we train it on image patches; it doesn't work on whole high-resolution images (1024x2018).
- For OpenGAN ^{fea}, we "synthesize" the RGB patches analytically –
    - for a synthesized feature, find the closest pixel-feature in the train-set and use the corresponding RGB patch as the "synthesized patch".
- OpenGAN ^{pix} synthesizes realistic patches in terms of color and tone, but notably underperforms OpenGAN ^{fea} (0.549 vs. 0.709 AUROC), which "synthesizes" ignored objects.

Thank you!

- Go beyond MNIST for the open-set research!
- Embrace outlier data in the real open world!
- Synthesize outliers to better span the open space!

Code: https://github.com/aimerykong/OpenGAN