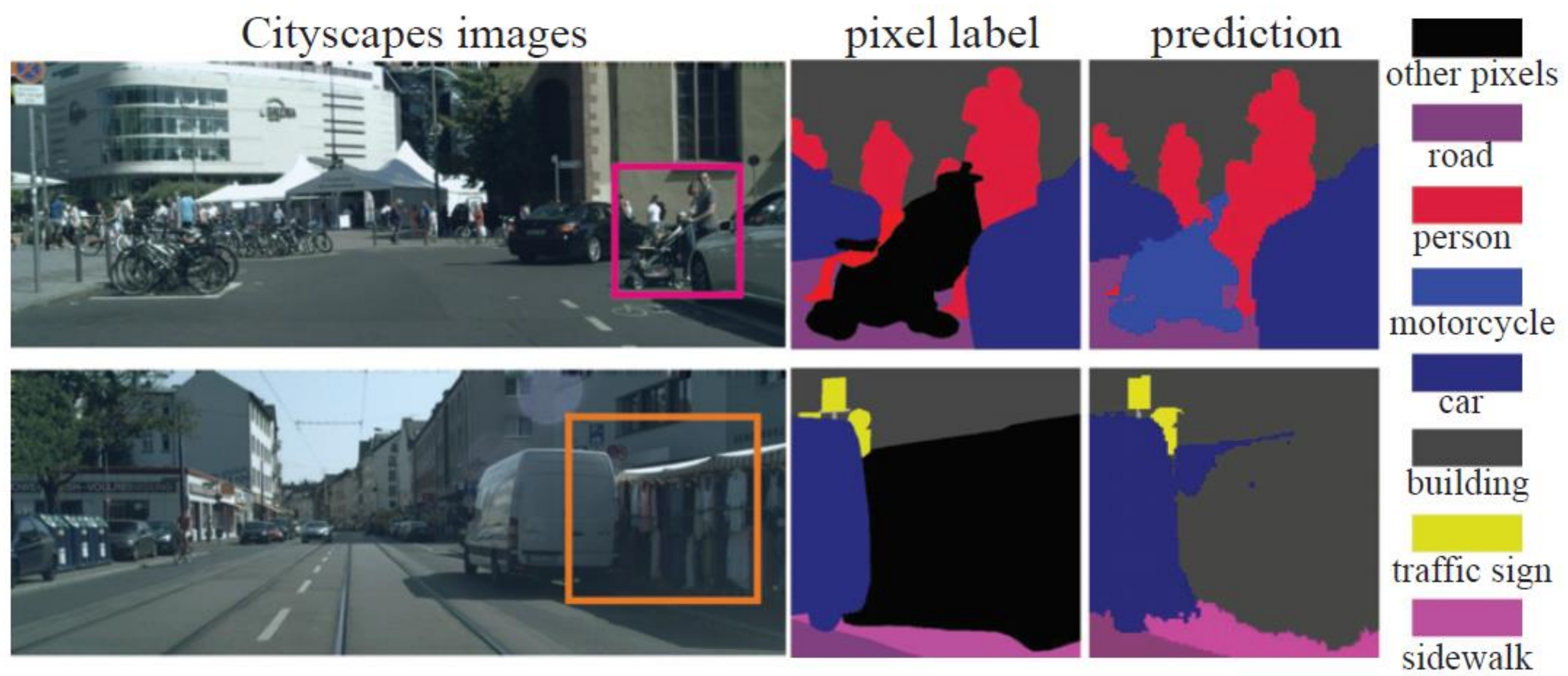
OpenGAN: Open-Set Recognition via Open Data Generation

Shu Kong*, Deva Ramanan*†

*Carnegie Mellon University, †Argo AI

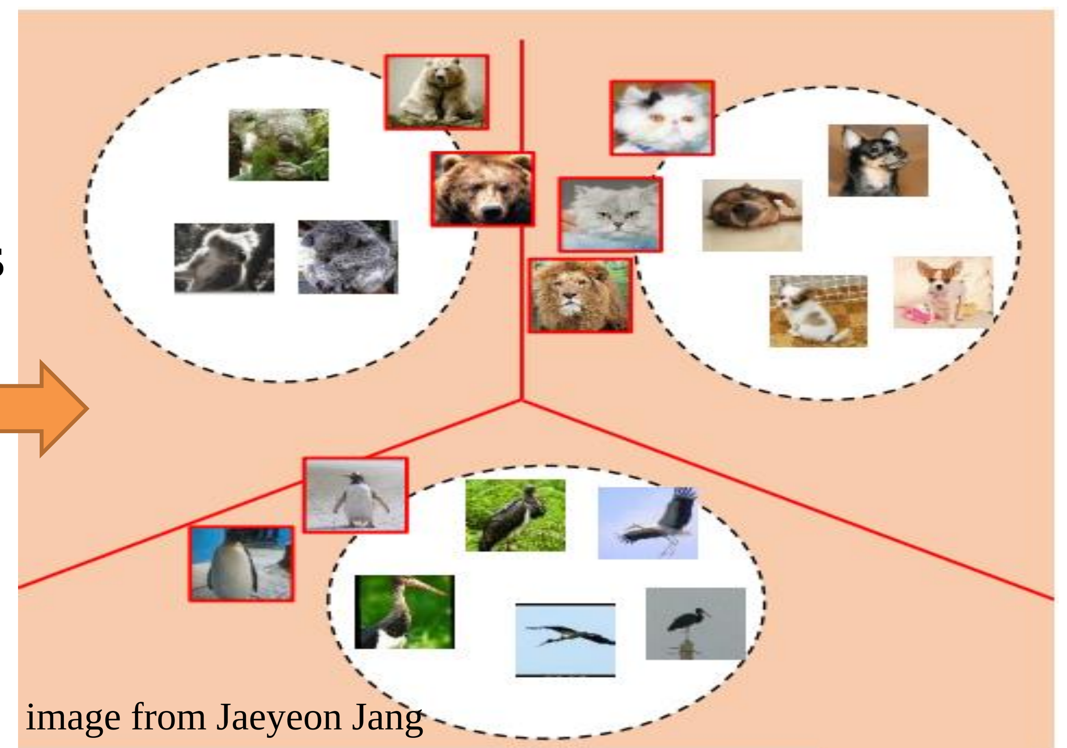
Motivation. Real-world machine learning systems need to analyze novel testing data that differs from the training data. Failing to recognize *unknown* objects causes serious safety concerns in autonomous vehicles (AVs).

A state-of-the-art semantic segmentation network has not been trained to recognize strollers or street-market (mid coln below). It misclassifies them as motorcycle and building (right coln below). Such misclassifications can be a critical mistake when fed into AVs because these objects require different plans for obstacle avoidance.



Problem formulation. Through the lens of K-way classification, a system should flag *unknown objects* not belonging to the pre-defined K classes, in addition to K-way classification. This is crisply formulated as *Open-Set Recognition*.

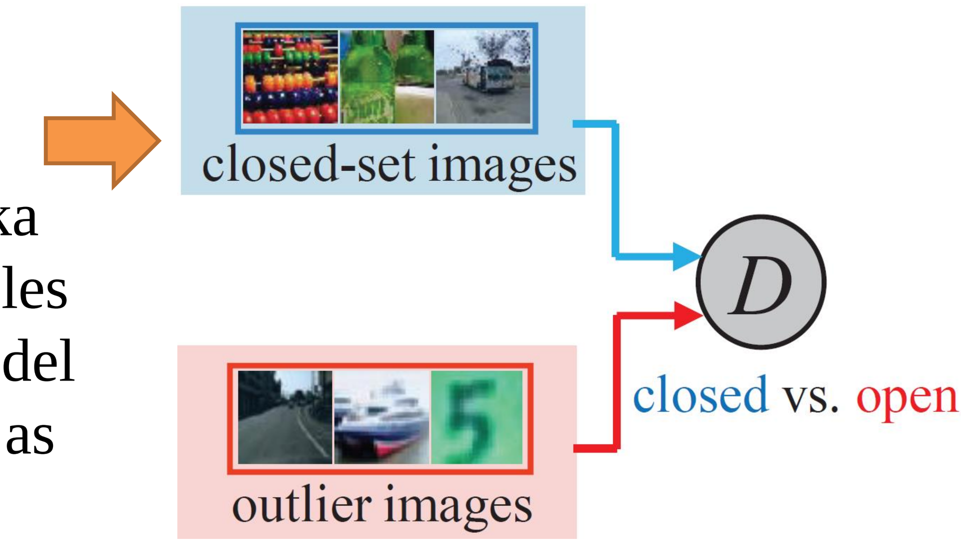
Open-Set Recognition requires recognizing examples from the pre-defined K classes and identifies *unknown examples* that belong to some other classes outside the K classes.



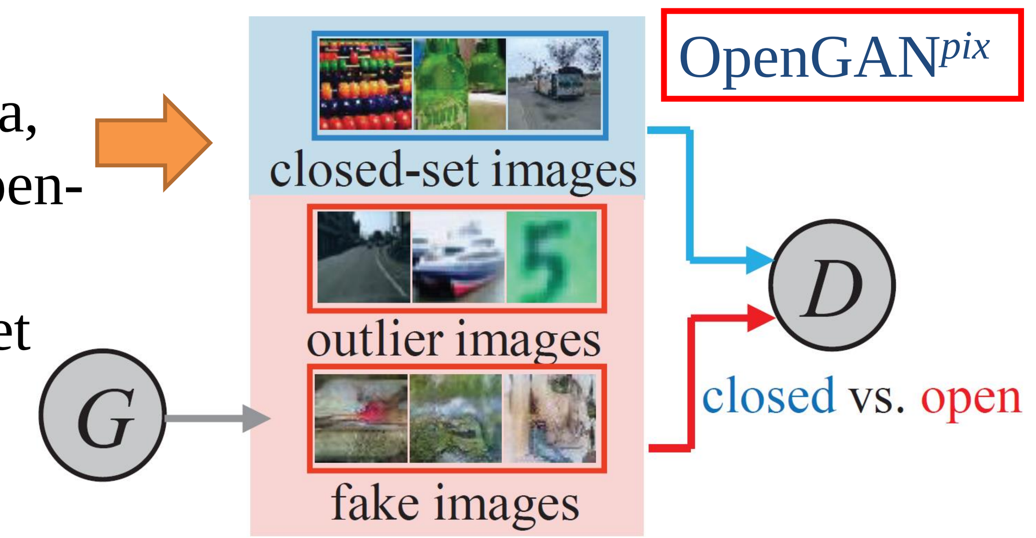
Our solution: OpenGAN trains a post-hoc open-set classifier to identify the open-set over features extracted from the off-the-shelf K-way network. The classifier is *adversarially* trained on real outlier data and fake data. To generate the fake data, we train a GAN by fooling the open-set classifier. That said, open-set classifier ≡ GAN-discriminator !

Key insights in OpenGAN

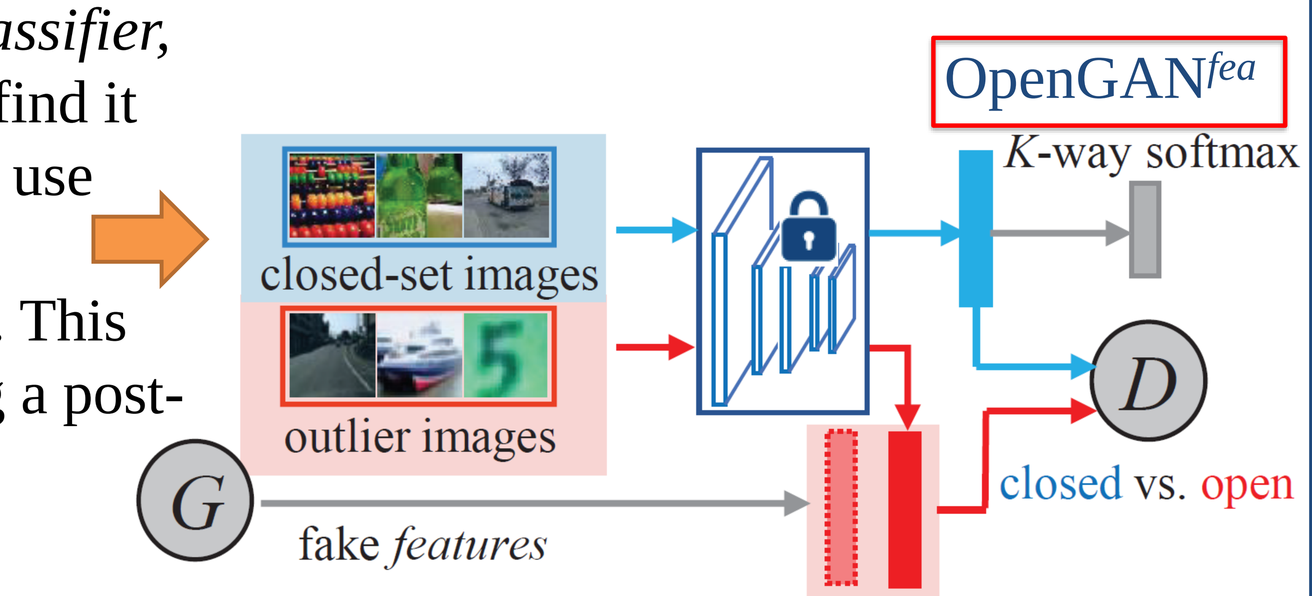
(a) *Use outlier data* for training the *discriminative* open-set classifier (aka Outlier Exposure). Outlier data enables stable training and validation for model selection. But outliers can be biased as they will not span the open world.



(b) *Use fake data to augment the outlier data.* To generate fake data, we train a GAN by fooling the open-set classifier. In this sense, the GAN-discriminator is the open-set classifier.



(c) *Using off-the-shelf features to train the open-set classifier, rather than pixels.* We find it much more effective to use features for open-set recognition than pixels. This leads to simply training a post-hoc lightweight model.



Experiment I: Cross-Dataset Open-Set Recognition

Metrics: Area Under ROC Curve and macro-F1 over (K+1) classes

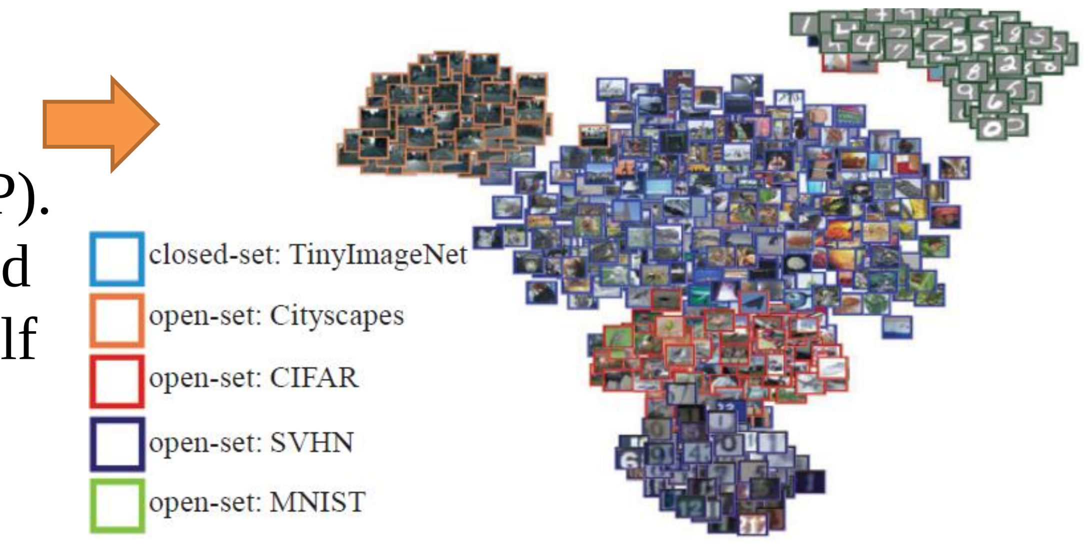
Setup: train a 200-way classification network on TinyImageNet train-set as the closed-set and another dataset (e.g., MNIST) as the outlier data, test on TinyImageNet test-set and a third dataset (e.g., CIFAR) as the open-set.

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}
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**

Conclusion

1) Using features is better than pixels, ref. OpenGAN^{fea} vs OpenGAN^{pix}
2) OpenGAN is better than binary classifiers, ref. OpenGAN^{fea} vs. CLS^{fea} which is a binary open-vs-closed classifier trained on features.
3) OpenGAN is better than (K+1) classifiers, ref. OpenGAN vs. (K+1)
4) discriminative is better than generative, ref. OpenGAN vs. GMM

Visualization. Recall that the open-set classifier is the GAN-discriminator (a nonlinear MLP). It effectively groups closed- and open-set data in the off-the-shelf feature space. This intuitively shows why OpenGAN works.



Experiment II: Open-Set Discrimination (toyish but "standard")

Setup: split a dataset (e.g., MNIST) into a closed-set (e.g., 0-5 digits) and an open-set (e.g., 6-9 digits) w.r.t class labels; train only on closed-set but test on both closed- and open-set.

Conclusion holds as in Experiment I.

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**

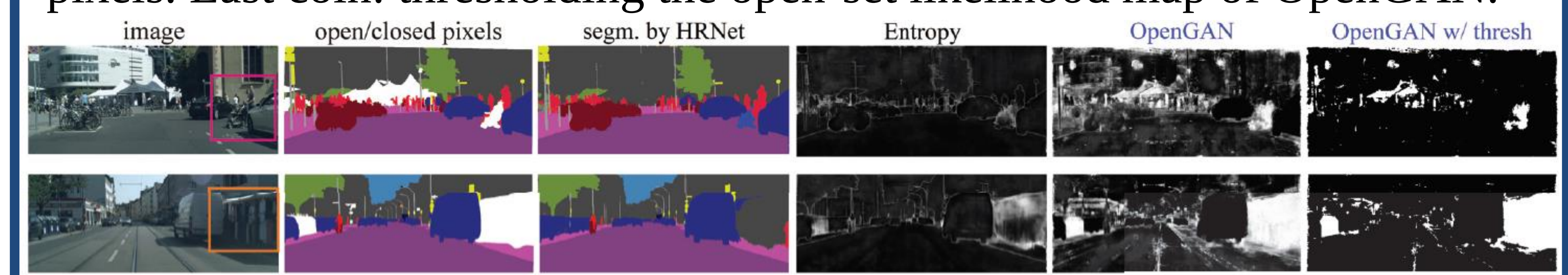
Experiment III: Open-Set Semantic Segmentation

Setup: recognize unknown object pixels in the context of semantic segmentation.

Conclusion holds as in Experiment I. Importantly, OpenGAN performs significantly better than existing open-set methods. Moreover, unlike OpenGAN which is a simple discriminative method, image-reconstruction based methods (not shown here) do not work well because of the difficulties in reconstructing high-resolution images.

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**

Visualization. Second coln: white pixels are the "ground-truth" open-set pixels. Last coln: thresholding the open-set likelihood map of OpenGAN.



Acknowledgements. This work was supported by the CMU Argo AI Center for Autonomous Vehicle Research.