Background Data Resampling for Outlier-Aware Classification

Yi Li Nuno Vasconcelos

UC San Diego

SVCL

Out-of-distribution Detection

- Deep neural nets tend to produce overconfident predictions, specifically on
  - Misclassified examples (Guo et al., 2017)
  - Inputs that do not belong to any training class (Bendale et al., 2016)

- Out-of-distribution (OOD) detection: Discriminate outliers from regular test data
  - i.e. identify samples from different prob. distribution than training set
  - Existing methods: Input preprocessing (Liang et al., 2018), additional loss functions (Lee et al., 2018)
  - Auxiliary background data effective (Hendrycks et al., 2019), less explored

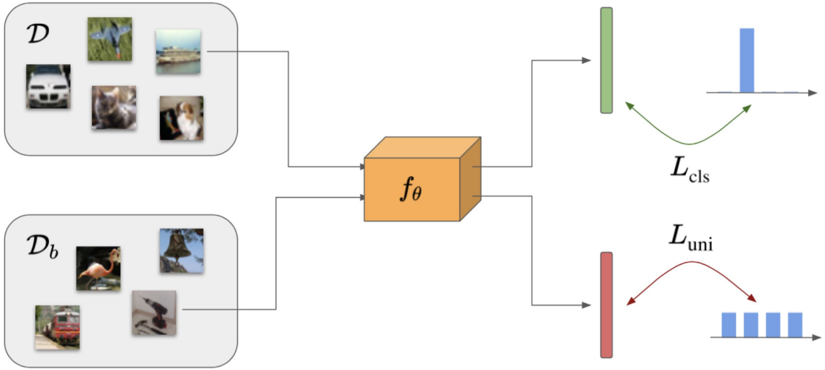
SVCL

OOD Detection with Background Data

- Formulation: Two objectives

$$L(\theta; \mathcal{D}, \mathcal{D}_b) = L_{\text{cls}}(\theta; \mathcal{D}) + \alpha L_{\text{uni}}(\theta; \mathcal{D}_b)$$

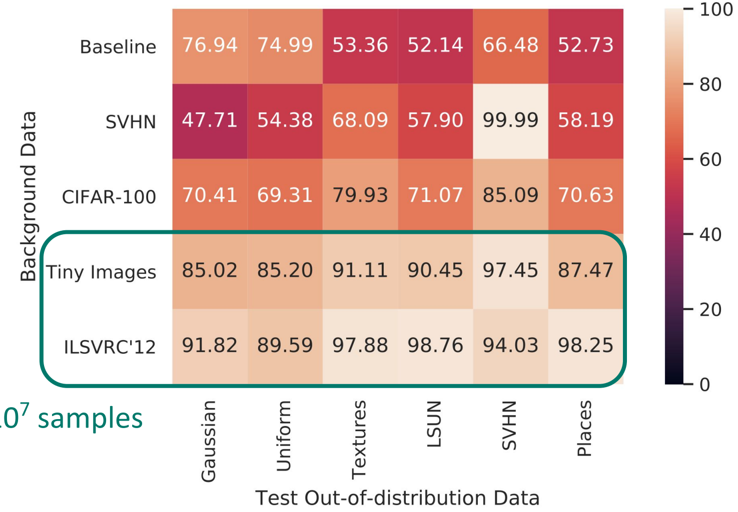
  - L_cls --- Classify in-distribution samples w/ high confidence output
  - L_uni --- Detect out-of-distribution samples w/ low confidence output



- Challenge: Dataset size

OOD Detection with Background Data

- Large background dataset needed!
    - Additional storage & training time
    - Trade-off between detection quality & sample size



10^6 -- 10^7 samples

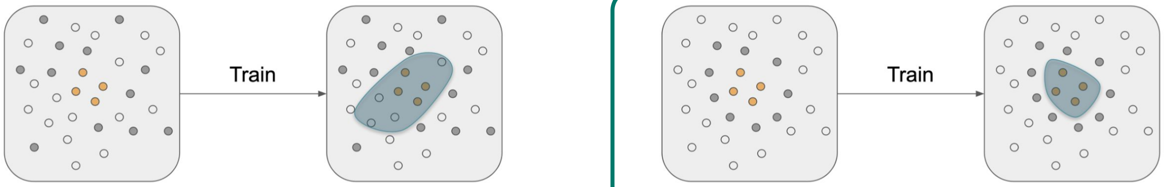
SVCL

OOD Detection with Background Data

- What background data to use?

(a) Small background dataset: Efficient but inaccurate.

(b) Large background dataset: Accurate but inefficient.

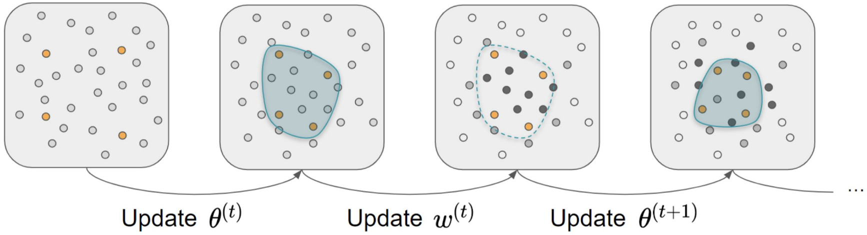


(c) Uniformly resampled dataset: Efficient but inaccurate.

(d) Adversarially resampled dataset: Accurate and efficient.

Background Data Resampling

- Intuition
  - Assign individual weights to background samples
  - Adversarially update sample weights & classifier parameters
  - Use optimized weights to sample background subset



SVCL

Background Data Resampling

- Example reweighting
  - Assign $w_i \geq 0$ to sample x_i in background dataset $\mathcal{D}_b$
  - Reweight training loss using w_i
  - Special case when $w_i \in \{0, 1\}$ Reweighted loss = Loss on background subset $\mathcal{D}'_b$

$$
\begin{aligned}
L_{\text{out}}(\theta; w) &= \frac{1}{|\mathcal{D}'_b|} \sum_{(x,y) \in \mathcal{D}'_b} L_{\text{uni}}(f(x;\theta)) \\
&= \frac{1}{\sum_i w_i} \sum_{i=1}^{|\mathcal{D}_b|} w_i L_{\text{uni}}(f(x_i;\theta)).
\end{aligned}
$$

SVCL

Background Data Resampling

- Adversarial resampling
  - Classifier updated to minimize reweighted loss
  - Sample weights updated to maximize reweighted loss, selecting the most challenging examples near the boundary of training distribution
- Background subset obtained through sampling w/ probability proportional to learned weights

---

Algorithm 1: Adversarial resampling, batch version.

Input: ID dataset $\mathcal{D}$, background dataset $\mathcal{D}_b$, pre-trained classifier θ, learning rate η_θ, η_w, loss coefficient α, total iterations T

Initialize: $w^{(0)} \leftarrow [1, \ \ldots, \ 1]$, $\theta^{(0)} \leftarrow \theta$;

for $t = 0, \ldots, T-1$ **do**

  Compute ID loss $l_{\text{in}}^{(t)} \leftarrow L_{\text{in}}(\theta^{(t)}; \mathcal{D})$;

  Compute OOD loss $l_{\text{out}}^{(t)} \leftarrow L_{\text{out}}(\theta^{(t)}; w^{(t)})$;

  Update classifier

  $$\theta^{(t+1)} \leftarrow \theta^{(t)} - \eta_\theta \nabla_{\theta^{(t)}} \left(l_{\text{in}}^{(t)} + \alpha l_{\text{out}}^{(t)} \right);$$

  Update weights

  $$w^{(t+1)} \leftarrow w^{(t)} + \eta_w \nabla_{w^{(t)}} \, l_{\text{out}}^{(t)};$$

Output: Resampling weights $w^{(T)}$.

SVCL

Experiments

- OOD detection performance
  - Training with background data improves OOD detection by large margin
  - Random sampling 10% of background samples hurt detection quality
  - Adversarial sampling gives similar or better performance, thanks to emphasis on samples near the boundary

Background $\mathcal{D}_b$	FPR95 ↓	AUROC ↑	AUPR ↑
None [13], $\gamma = 0$	31.45	90.72	62.77
Full, $\gamma = 100\%$	2.21	**99.41**	**95.06**
Random, $\gamma = 10\%$	2.85	99.14	92.92
Resampled, $\gamma = 10\%$	**1.94**	99.37	94.16

(a) In-distribution $\mathcal{D} =$ CIFAR-10.

Background $\mathcal{D}_b$	FPR95 ↓	AUROC ↑	AUPR ↑
None [13], $\gamma = 0$	54.81	76.71	33.98
Full, $\gamma = 100\%$	8.51	97.03	81.16
Random, $\gamma = 10\%$	11.08	96.08	76.17
Resampled, $\gamma = 10\%$	**6.40**	**97.76**	**83.75**

(b) In-distribution $\mathcal{D} =$ CIFAR-100.

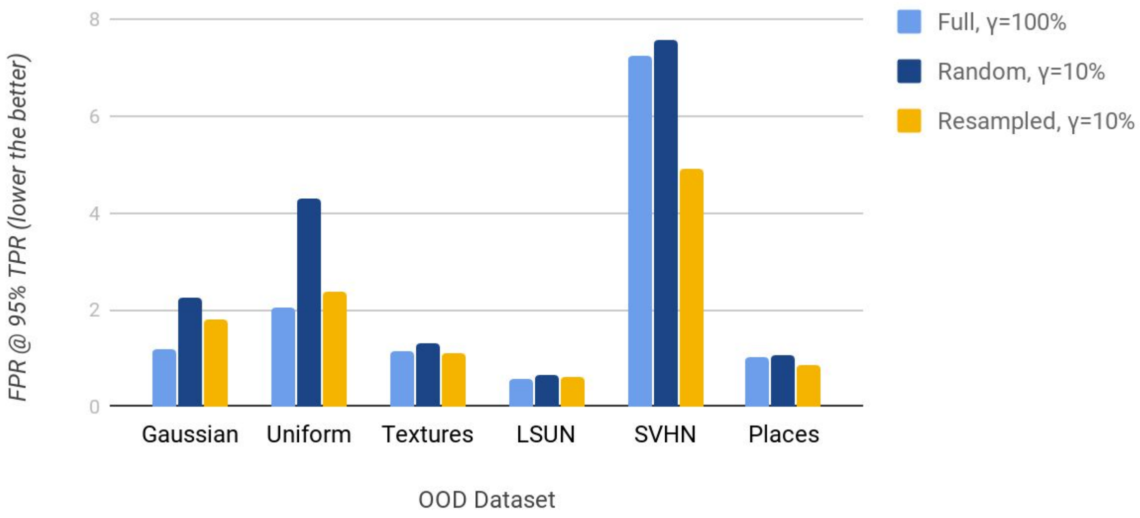
Background $\mathcal{D}_b$	FPR95 ↓	AUROC ↑	AUPR ↑
None [13], $\gamma = 0$	62.41	72.01	30.73
Full, $\gamma = 100\%$	3.77	99.39	97.70
Random, $\gamma = 10\%$	8.17	98.19	95.22
Resampled, $\gamma = 10\%$	**1.25**	**99.64**	**98.86**

(c) In-distribution $\mathcal{D} =$ Tiny ImageNet.

SVCL

Experiments

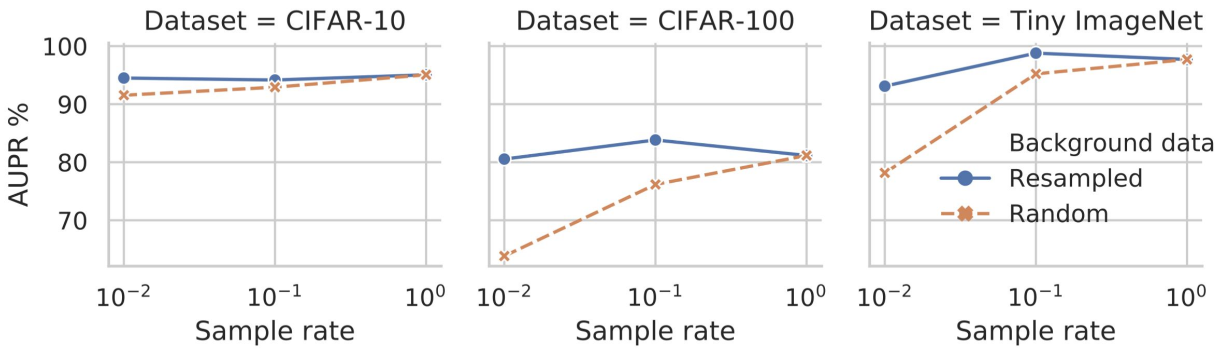
- OOD detection performance: Breakdown by OOD test sets (In-distribution: CIFAR-10)



SVCL

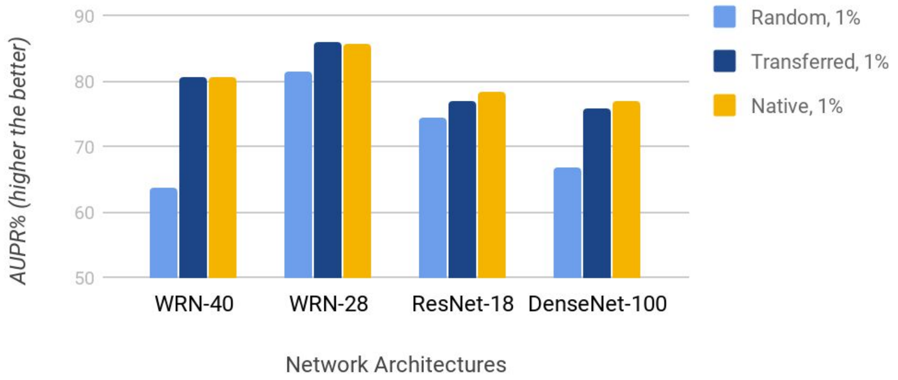
Experiments

- How many background samples to use?
  - Detection quality vs. Sample rate (% of background data used)



Experiments

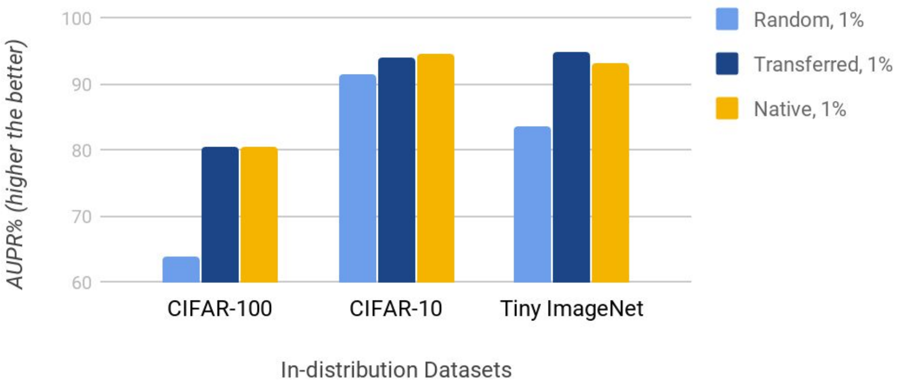
- Does resampled background data work under different training settings?
  - Generalization across models
  - Generalization across in-distribution datasets



SVCL

Experiments

- Does resampled background data work under different training settings?
    - Generalization across models
    - Generalization across in-distribution datasets



SVCL

Conclusions

- Motivations
  - Background data for training OOD detection
  - Trade-off between sample size and detection quality
- Background data resampling
  - Reweight background samples
  - Adversarially updating sample weights & classifier
- Results
  - Training with resampled dataset > random sample of equal size, sometimes outperforming full background data
  - Improvement is consistent at different resampling rates
  - Resampled data generalizes in different training settings

SVCL