模式识别与神经计算研究组

PAttern Recognition and NEural Computing

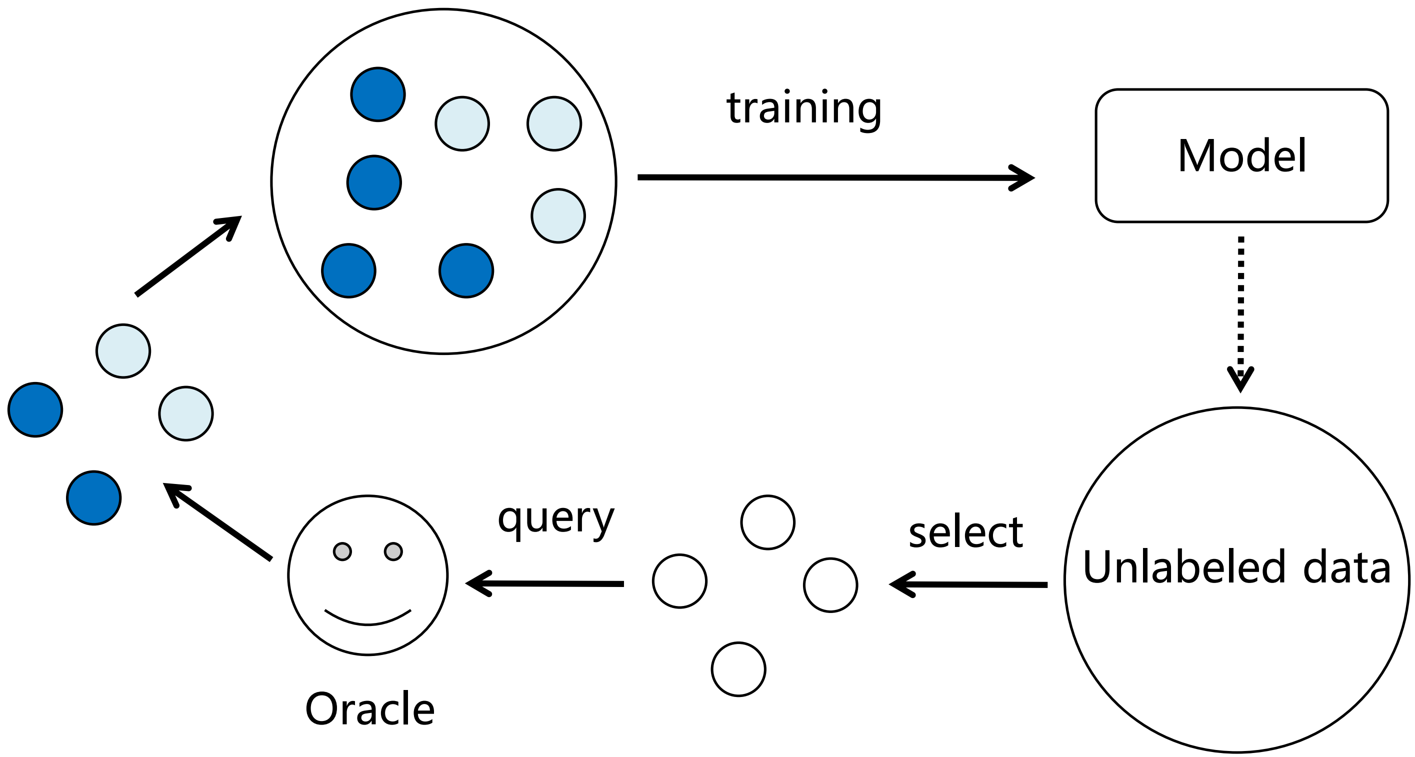
Active Image Synthesis for Efficient Labeling

Jialei Chen, Yujia Xie, Kan Wang, Chuck Zhang, Mani A. Vannan, Ben Wang, and Zhen Qian

Georgia Institute of Technology

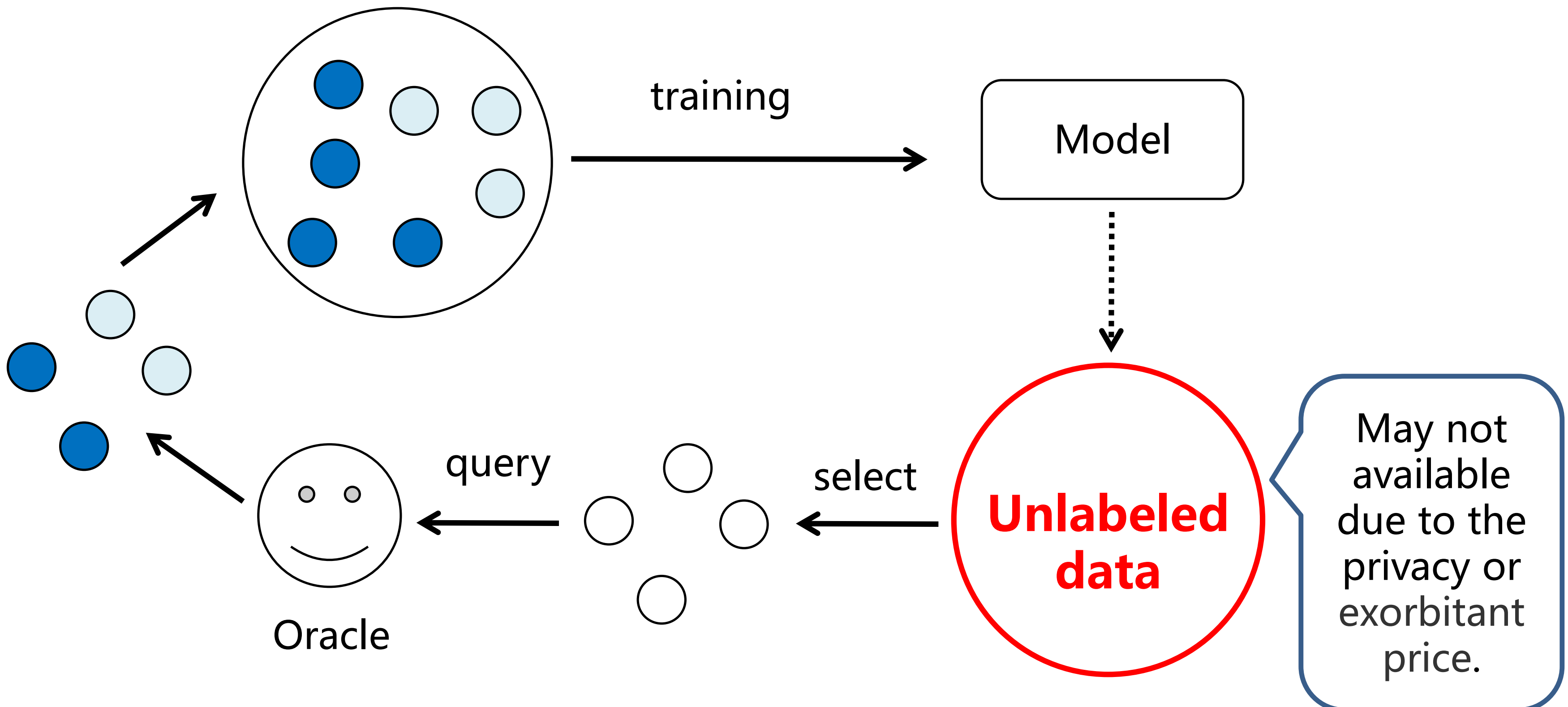
Hippocrates Research Lab, Tencent America

TPAMI 2021

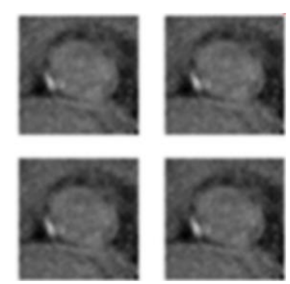


Goal: query less for more.

Introduction



A hospital may only have around **a hundred** historical records of pre-surgical CT scans of aortic stenosis **over the years.**



Main Contributions

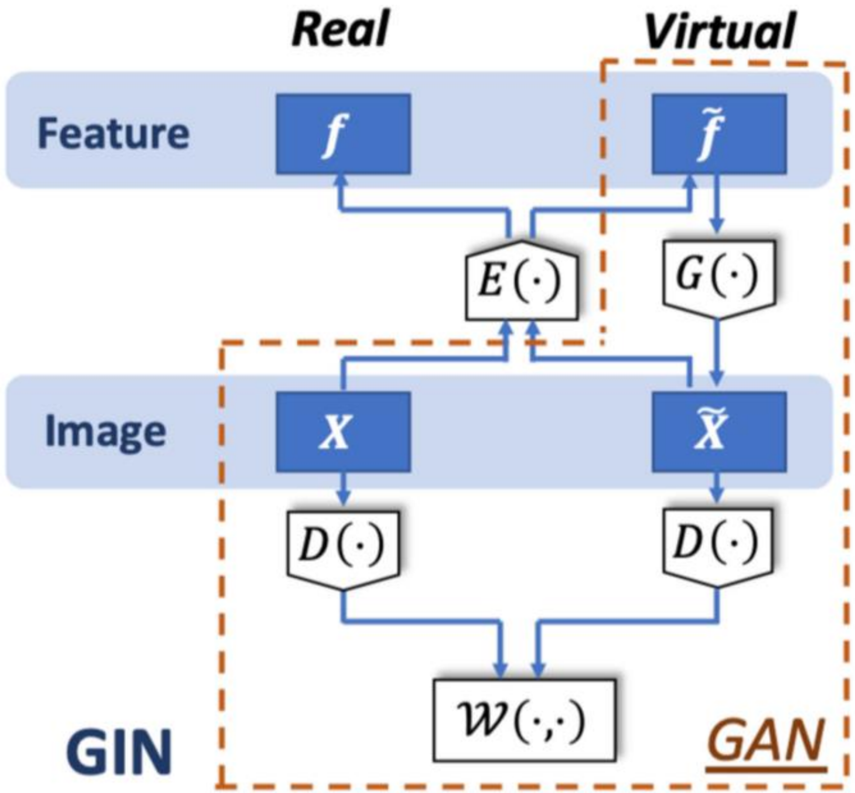
- Propose a generative invertible network (GIN) to generate images with small training data.
- Propose an active sampling method to select informative features from the embedded feature space.
- Conduct experiment on a real aortic stenosis dataset to demonstrate the superiority.

- **Further learn a encoder to inversely map the images to the latent feature vectors (for AL).**

$$\min_{E(\cdot)} \mathbb{E}_{\mathbf{X}\sim\mathcal{X}_n} ||E(\mathbf{X}) - f||_2^2,$$

However, the difficulty is that the feature f for the actual image X is unknown

$$E(\cdot) = \underset{E(\cdot)}{\operatorname{argmin}}\, \mathbb{E}_{u\sim\mathcal{U}} ||E(G(u)) - u||_2^2,$$



Theorem 1. *Denote the target distribution measuring as $\mathcal{X}$ on image space $\{\mathbb{X}, \mathcal{B}[\mathbb{X}]\}$. Assume the generator $G(\cdot)$ is obtained by (3) with the training error $< \epsilon$ and encoder $E(\cdot)$ is obtained by (5) with the training error $< \delta$. If both $G(\cdot)$ and $E(\cdot)$ are Lipschitz-L continues, then the reconstruction error $\mathbb{E}_{x \sim \mathcal{X}} [G(E(x)) - x]^2$ can be bounded by $(L^2 + L + 1)\epsilon + L\delta$.*

$$\min_{G(\cdot)} \max_{D(\cdot)} \mathbb{E}_{x \sim \mathcal{X}_n}[D(x)] - \mathbb{E}_{u \sim \mathcal{U}}[D(G(u))], \tag{3}$$

$$E(\cdot) = \operatorname*{argmin}_{E(\cdot)} \mathbb{E}_{u \sim \mathcal{U}} \|E(G(u)) - u\|_2^2, \tag{5}$$

Active sampling

- **Notations**

for any $f_0 \in \mathbb{F}$ where $\mathbb{F} = [-1, 1]^r$ is the feature space.

For any input image $\mathbf{X}_0 \in \mathbb{X}$, let $y_0 = C(\mathbf{X}_0)$ where $C(\cdot) : \mathbb{X} \mapsto [0, 1]^K$ is the classifier trained with the limited real data

- **Entropy definition**

$$H(\mathbf{X}_0) = -\sum_{k=1}^{K} y_0[k] \log (y_0[k]),$$

- **Entropy for synthesis images**

$$h(f_0) = H(G(f_0)) = -\sum_{k=1}^{K} E(G(f_0))[k] \log (E(G(f_0))[k]). \quad (7)$$

Active sampling

- **Select uncertain features**

$$f'_{1:m} = \operatorname*{argmin}_{f'_{1:m}} dist(\mathcal{F}'_m, \mu_h).$$

where $f'_{1:m} = \{f'_i\}_{i=1}^m$

The entropy $h(\cdot) : \mathbb{F} \mapsto [0, \log K]$ can also be viewed as a (unnormalized) probability density on the measurable space $\{\mathbb{F}, \mathcal{B}[\mathbb{F}]\}$. We denote this *uncertainty measure* as μ_h.

- **Take the energy distance**

$$D^2(F,G) = 2\,\mathrm{E}\,\|X - Y\| - \mathrm{E}\,\|X - X'\| - \mathrm{E}\,\|Y - Y'\| \geq 0,$$

$$\min_{f'_{1:m}} \sum_{i=1}^{m} \mathbb{E}_{\gamma \sim \mu_h} \|f'_i - \gamma\|_2 - \frac{1}{2m} \sum_{i=1}^{m} \sum_{j=1}^{m} \|f'_i - f'_j\|_2.$$

Active sampling

- **Incorporate the real images**

$$\min_{f'_{1:m}} \sum_{i=1}^{m} \mathbb{E}_{\gamma \sim \mu_h} \| f'_i - \gamma \|_2 - \sum_{i=1}^{m+n} \sum_{j=1}^{m+n} \frac{\| f'_i - f'_j \|_2}{2(m+n)}, \tag{10}$$

actual images with indies $i = m+1, \ldots, m+n$.

- **Generate images with $G(f')$**

- **labeling by an oracle**

Overall framework

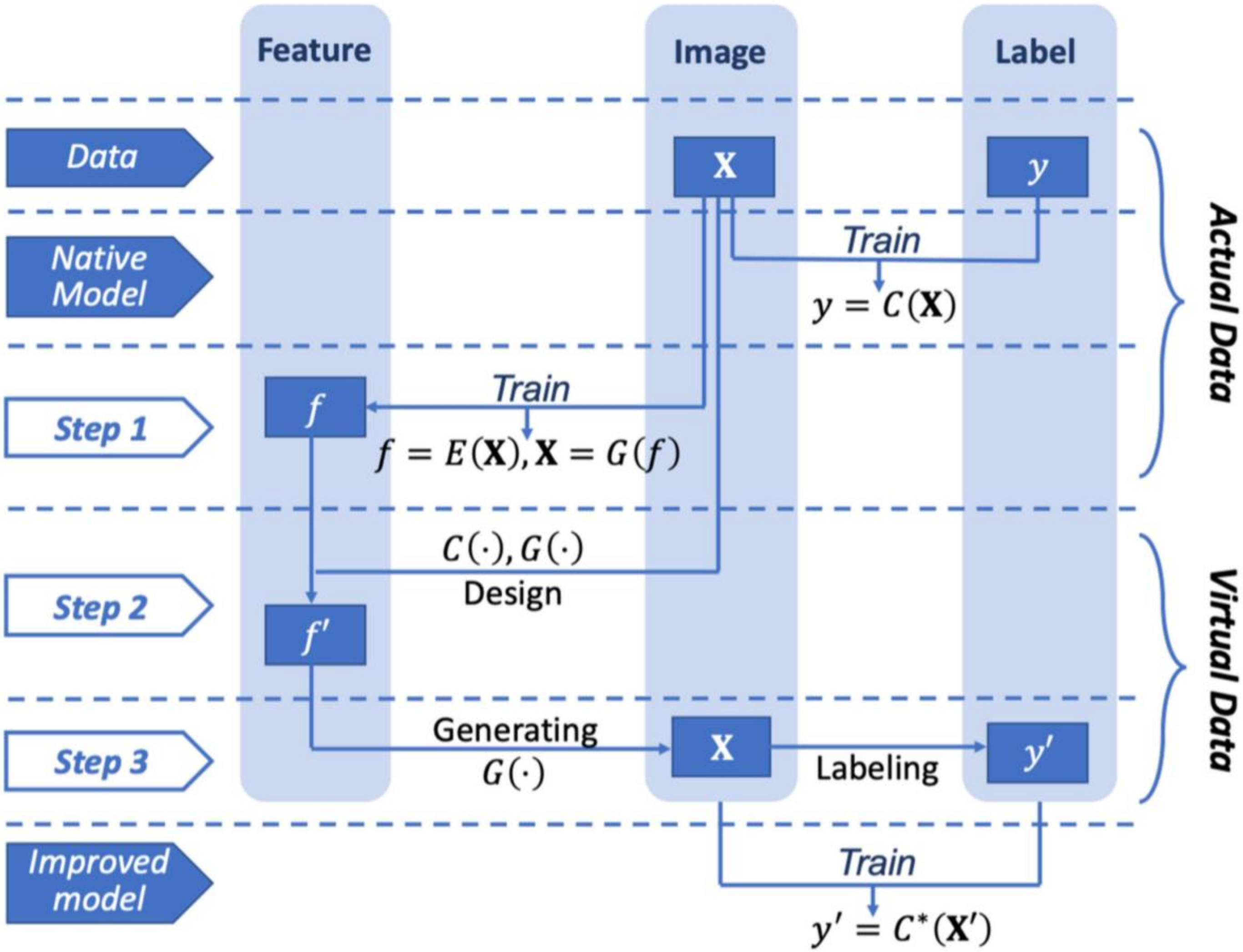
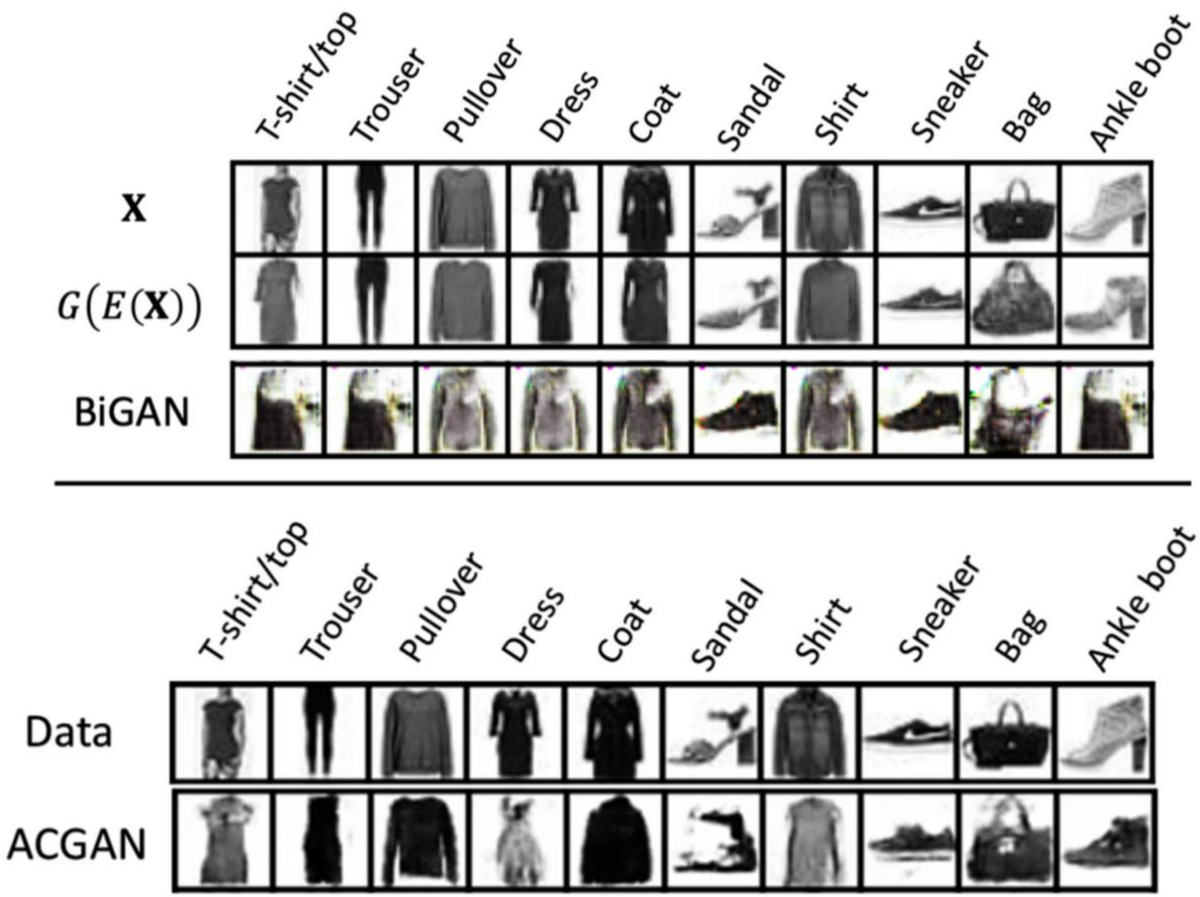


Fig. 2. The proposed three-step framework AISEL to efficiently sample AISEL dataset and improve classification.

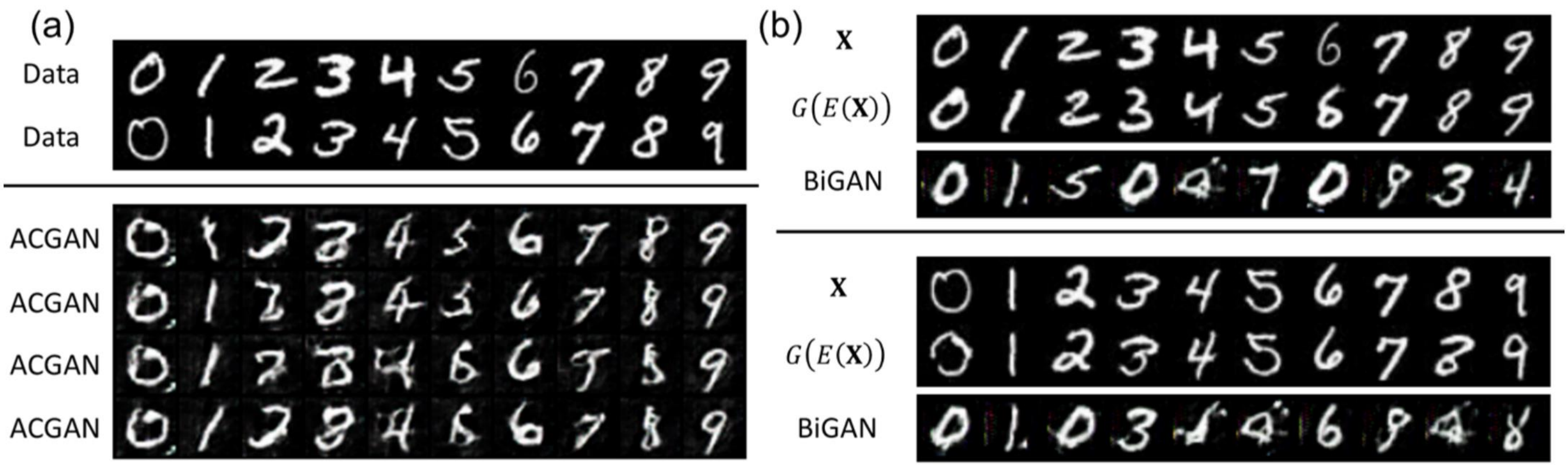
- **Toy Computer Vision Applications (Fashion and MNIST datasets). Visualization of image synthesis**

- 400 data is available for training

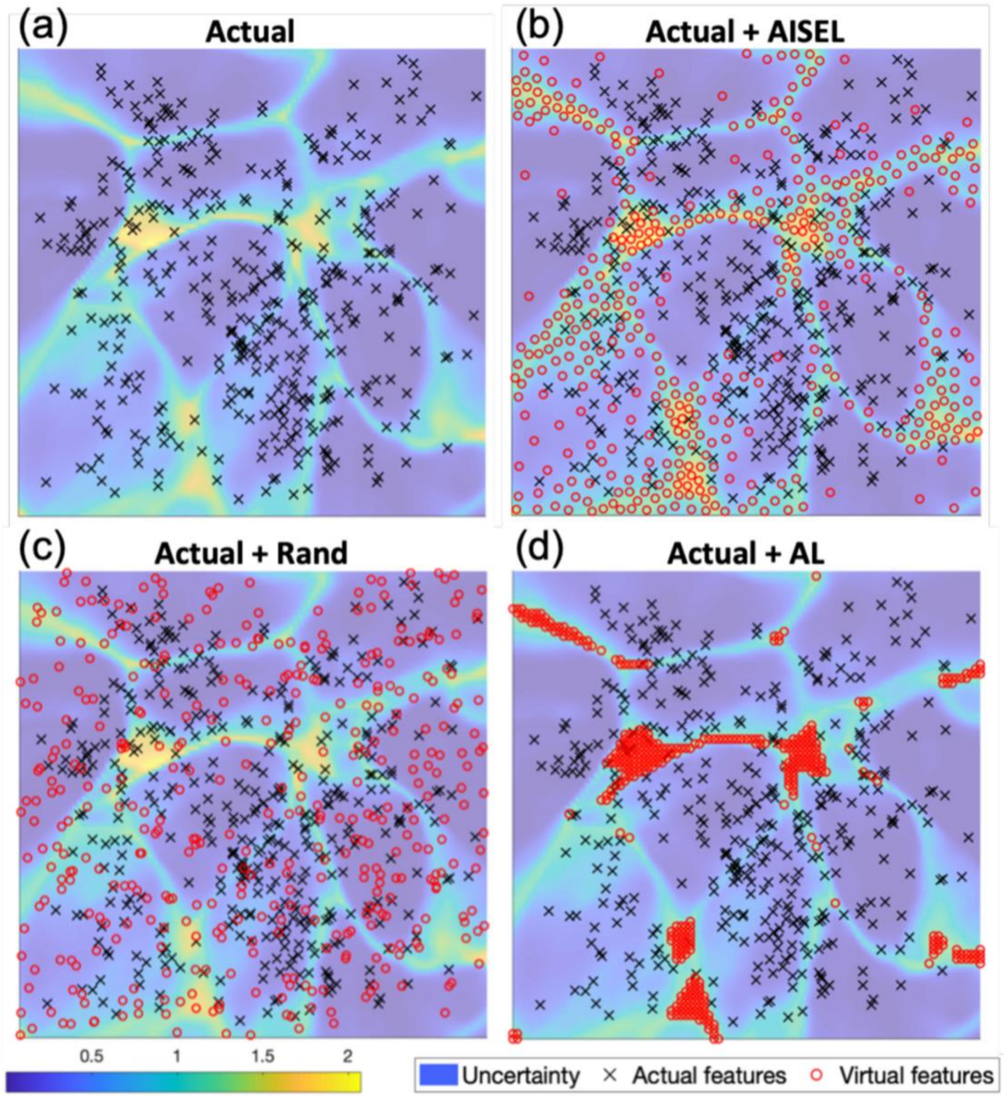


Experiments

- **Visualization of image synthesis**



- **Visualization of active sampling (Fashion dataset)**



Note that the **labels are obtained by the oracle model** (using all 60,000 training data and WideResNet architecture)

AL method select data by uncertainty.

- **The Classification Accuracy on Fashion and MNIST**

Finetune pretrained ResNet18 (ImageNet)

Augment training data with generated images

Random sampling features for image synthesis

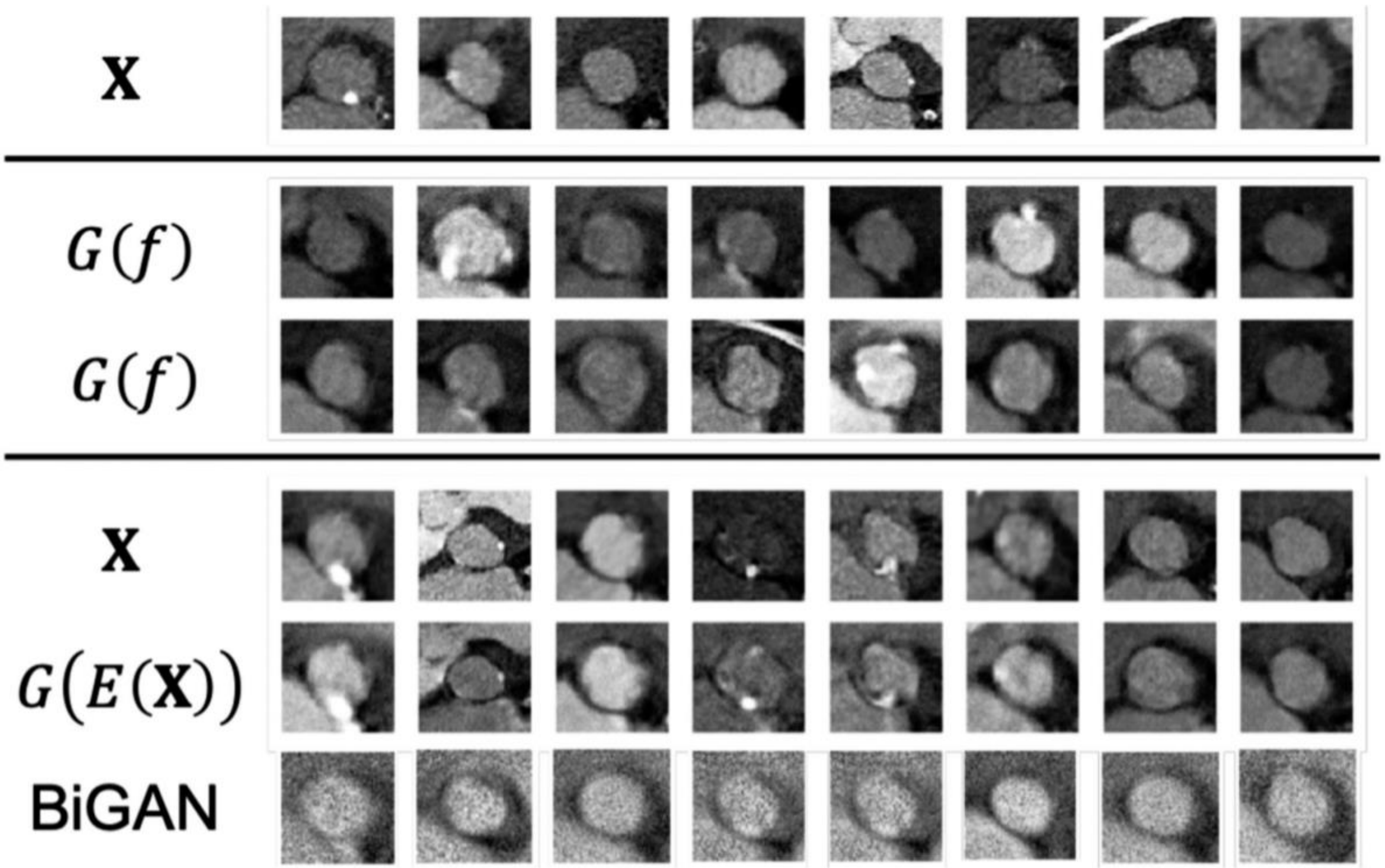
	Native (400)	Transfer	ACGAN	Rand (+400)	Rand (+5000)
Fashion	**72.8%**	74.3%	72.3%	76.4%	80.7%
MNIST	**88.2%**	87.4%	85.9%	90.2%	91.6%

Query by uncertainty

Use real images

	AISEL (+400)	AL (+400)	Oracle (800)	Oracle (all)
Fashion	**81.9%**	78.2%	81.3%	96.7%
MNIST	**91.2%**	90.4%	91.3%	99.2%

- **The visualization on aortic stenosis dataset**
  - 168 CT scan data. 75% for training.



- **The Classification Accuracy on aortic stenosis dataset**

	Native Model (126)				Randomly Generated (+1134)			
Fold	Accu.	Sens.	Spec.	F1	Accu.	Sens.	Spec.	F1
1	64.29	52.17	78.95	61.54	69.05	60.87	78.95	68.29
2	56.96	47.26	66.67	48.65	64.29	61.90	66.67	60.00
3	57.14	50.00	63.64	52.63	71.43	52.94	84.00	58.82
4	59.52	55.00	63.64	56.41	64.29	60.00	68.18	60.00
Ave	**59.48**	51.11	68.23	54.81	**67.27**	58.93	74.45	61.78

	AISEL (+1134)				Randomly Generated (+10000)			
Fold	Accu.	Sens.	Spec.	F1	Accu.	Sens.	Spec.	F1
1	76.19	73.91	78.95	74.42	78.57	78.26	78.95	77.27
2	73.81	76.19	71.43	71.43	76.19	71.43	80.95	73.17
3	80.95	76.47	84.00	76.92	85.71	82.35	88.00	82.05
4	71.43	75.00	68.18	69.77	73.81	70.00	77.27	70.00
Ave	**75.60**	75.39	75.64	73.14	**78.57**	75.51	81.29	75.62

Conclusion

- **This paper proposes an active image synthesis method to generate informative images.**

- **Propose a generative invertible network (GIN) to generate images with small training data, and map the real images to the latent features.**

- **Propose an active sampling method which considers both uncertainty and diversity (by energy distance) to select informative features for image generation.**

THANKS