SEMANTIC-AWARE KNOWLEDGE PRESERVATION FOR ZERO-SHOT SKETCH-BASED IMAGE RETRIEVAL

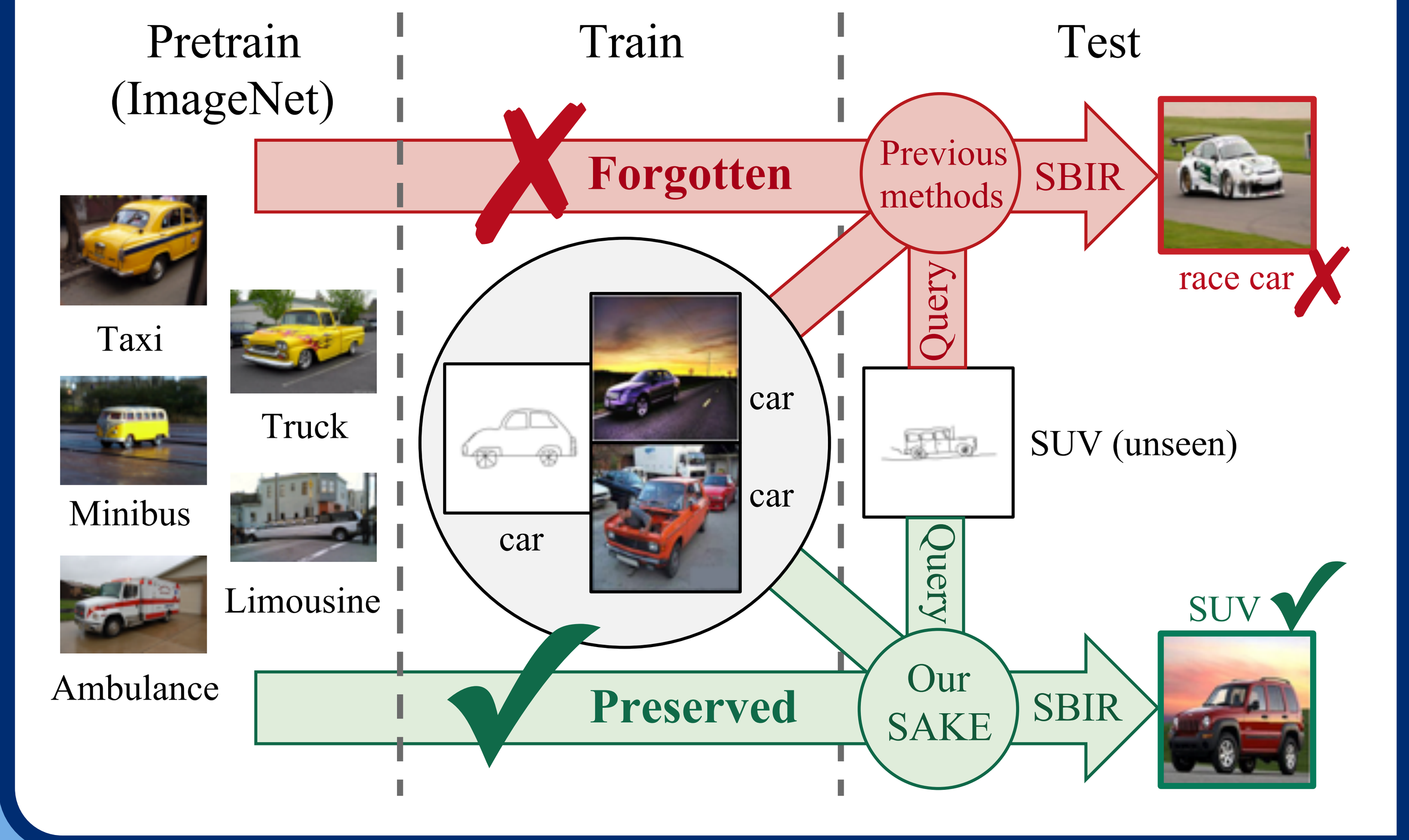
JOHNS HOPKINS UNIVERSITY   HUAWEI

Qing Liu[1] Lingxi Xie[1,2] Huiyu Wang[1] Alan L. Yuille[1]

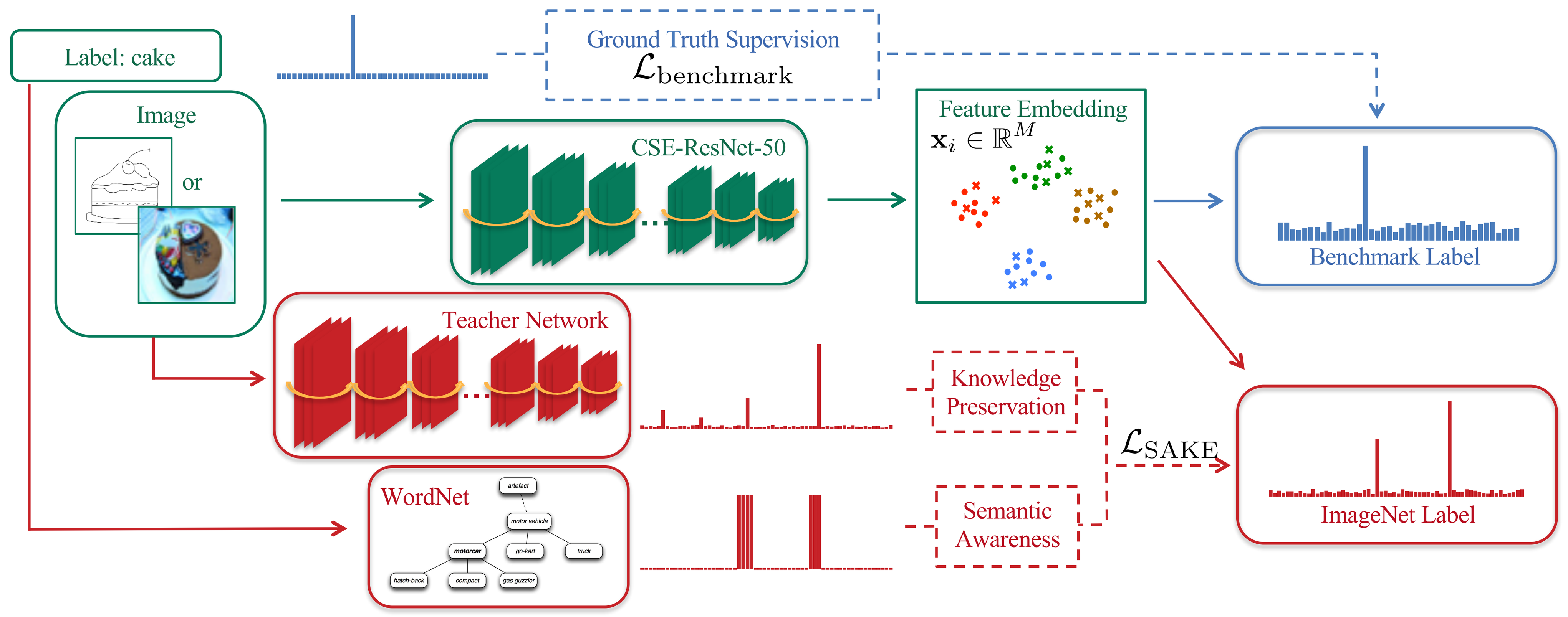
[1]Johns Hopkins University [2]Noah's Ark Lab, Huawei Inc.

https://github.com/qliu24/SAKE

MOTIVATION



FRAMEWORK



- Benchmark Loss

$$-\log \frac{\exp\left(\boldsymbol{\alpha}_{y_i}^\top \mathbf{x}_i + \beta_{y_i}\right)}{\sum_{k\in\mathcal{C}^S}\exp\left(\boldsymbol{\alpha}_k^\top \mathbf{x}_i + \beta_k\right)}$$

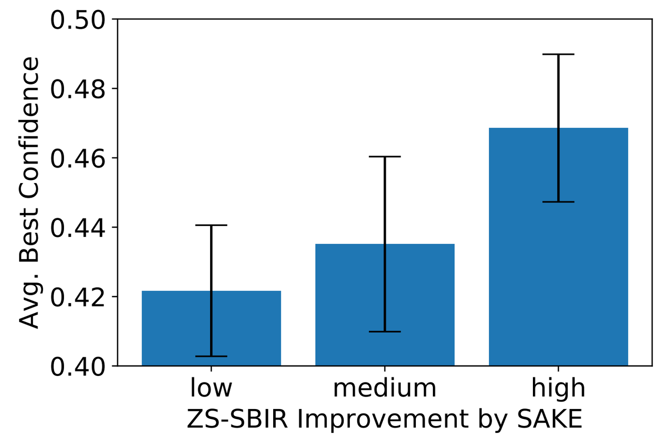
- Teacher Loss

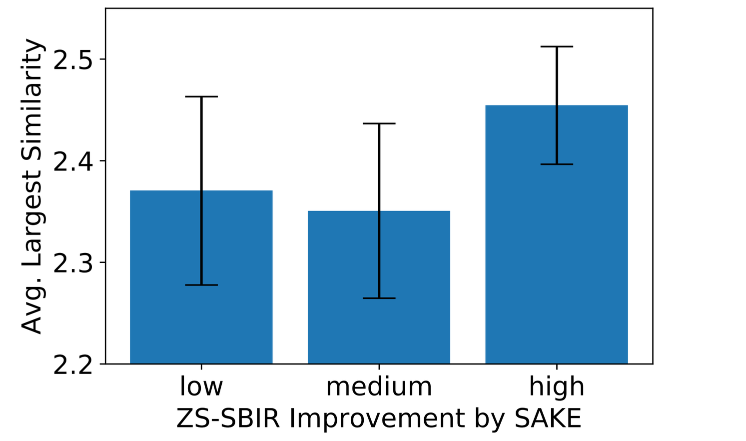
$$\sum_{m\in\mathcal{C}^O} -q^t_{i,m}\log \frac{\exp\left(\boldsymbol{\zeta}_m^\top \mathbf{x}_i + \eta_m\right)}{\sum_{l\in\mathcal{C}^O}\exp\left(\boldsymbol{\zeta}_l^\top \mathbf{x}_i + \eta_l\right)}$$

- SAKE Loss

$$\sum_{m\in\mathcal{C}^O} -q_{i,m}\log \frac{\exp\left(\boldsymbol{\zeta}_m^\top \mathbf{x}_i + \eta_m\right)}{\sum_{l\in\mathcal{C}^O}\exp\left(\boldsymbol{\zeta}_l^\top \mathbf{x}_i + \eta_l\right)}$$

MORE ANALYSIS & CONCLUSION





- We investigate the problem of ZS-SBIR from the viewpoint of domain adaptation.
- SAKE helps preserve previously acquired knowledge while fine-tuning a pre-trained model to improve the model's transfer ability.
- The performance gain of SAKE is mainly from the more properly structured feature embedding for photo images.

QUANTITATIVE RESULTS

- ZS-SBIR performance comparison of SAKE and existing methods.

Method	Dimension	TU-Berlin Ext. mAP@all	TU-Berlin Ext. Prec@100	Sketchy Ext. mAP@all	Sketchy Ext. Prec@100	Sketchy Ext. (Split2) mAP@200	Sketchy Ext. (Split2) Prec@200
ZSH [Yang et al., 2016]	64†	0.139	0.174	0.165	0.217	-	-
ZSIH [Shen et al., 2018]	64†	0.220	0.291	0.254	0.340	-	-
EMS [Lu et al., 2018]	512	0.259	0.369	-	-	-	-
	64†	0.165	0.252	-	-	-	-
CAAE [Yelamarthi et al., 2018]	4096	-	-	0.196	0.284	0.156	0.260
CVAE [Yelamarthi et al., 2018]	4096	-	-	-	-	0.225	0.333
SEM-PCYC [Dutta and Akata, 2019]	64	0.297	0.426	0.349	0.463	-	-
	64†	0.293	0.392	0.344	0.399	-	-
SAKE	512	**0.475**	**0.599**	**0.547**	**0.692**	**0.497**	**0.598**
	64†	**0.359**	**0.481**	**0.364**	**0.487**	**0.356**	**0.477**

- ZS-SBIR and ZS-PBIR mAP@all for different backbone models with different loss terms.

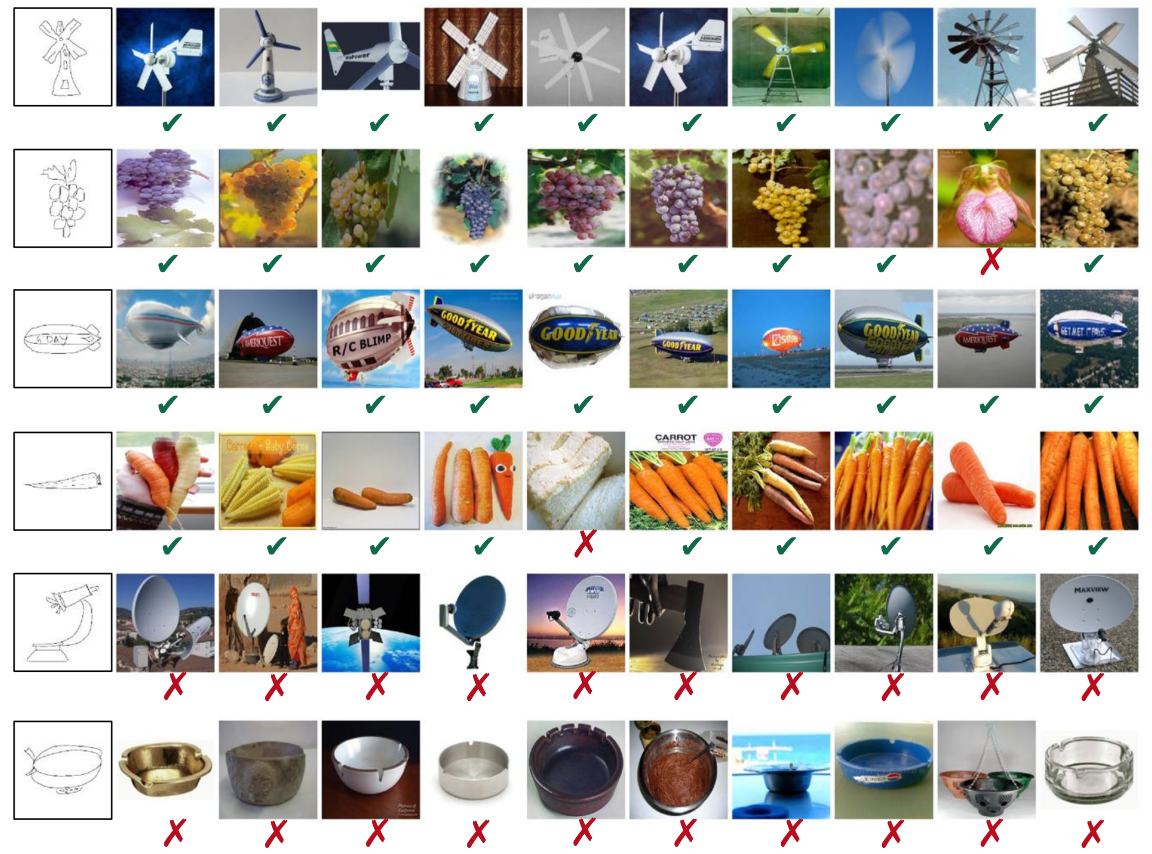
BackBone	ZS-SBIR pretrained	ZS-SBIR $\mathcal{L}_{\text{B}}$	ZS-SBIR $\mathcal{L}_{\text{B}}+\mathcal{L}_{\text{T}}$	ZS-SBIR $\mathcal{L}_{\text{B}}+\mathcal{L}_{\text{SAKE}}$	ZS-PBIR pretrained	ZS-PBIR $\mathcal{L}_{\text{B}}$	ZS-PBIR $\mathcal{L}_{\text{B}}+\mathcal{L}_{\text{T}}$	ZS-PBIR $\mathcal{L}_{\text{B}}+\mathcal{L}_{\text{SAKE}}$
AlexNet	0.074	0.267	0.275	0.275	0.386	0.393	0.427	0.432
ResNet-50	0.081	0.352	0.395	0.413	0.640	0.542	0.666	0.670
CSE-ResNet-50	0.068	0.353	0.426	**0.434**	0.635	0.558	0.673	**0.683**

ANALYSIS

- **Knowledge Preservation**

	ImageNet Acc@1 for Deep Feature Linear Classifiers pretrained	finetuned	SAKE
AlexNet	56.29	45.54	51.39
SE-ResNet-50	77.43	59.56	67.44

- **Retrieval Examples**



- **t-SNE**