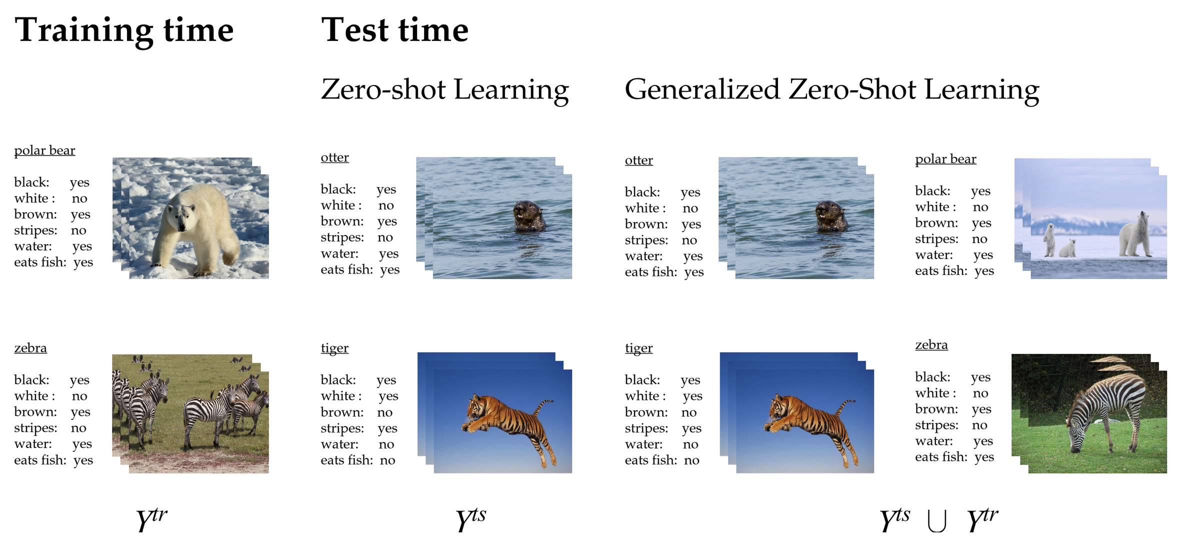
Zero-Shot Learning - The Good, the Bad and the Ugly

Yongqin Xian[1], Bernt Schiele[1], Zeynep Akata[1,2]

[1] Max-Planck Institute for Informatics [2] University of Amsterdam

Zero-Shot Evaluation Protocol, Evaluated Methods and Evaluation Metrics



Dataset	Size	$\lvert\mathcal{Y}\rvert$	$\lvert\mathcal{Y}^{tr}\rvert$	$\lvert\mathcal{Y}^{ts}\rvert$
SUN	$14K$	717	580 + 65	72
CUB	$11K$	200	100 + 50	50
AWA1	$30K$	50	27 + 13	10
AWA2 [11]	$37K$	50	27 + 13	10
aPY	$1.5K$	32	15 + 5	12

ImageNet Split	$\lvert\mathcal{Y}^{ts}\rvert$
ImageNet 21K - $\mathcal{Y}^{tr}$	20345
Within 2/3 hops from $\mathcal{Y}^{tr}$	1509/7678
Most populated classes	$500/1K/5K$
Least populated classes	$500/1K/5K$

The Good: Zero-Shot learning attracts lots of attention
The Bad: No agreed evaluation protocol
The Ugly: Test classes overlap ImageNet 1K

Group	Method	Main Idea
Linear compatibility	ALE [1]	Learn linear embedding by weighted ranking loss
	DEVISE [4]	Learn linear embedding by ranking loss
	SJE [2]	Learn linear embedding by multi-class SVM loss
	ESZSL [8]	Apply square loss and regularize the embedding
	SAE [5]	Learn linear embedding with autoencoder
Non-linear compatibility	LATEM [10]	Learn piece-wise linear embedding
	CMT [9]	Learn non-linear embedding by neural network
Two-stage inference	DAP [6]	Predict attributes → unseen class
	IAP [6]	Predict seen class → attributes → unseen class
	CONSE [7]	Predict seen class→embedding→ unseen class
Hybrid	SSE [12]	Learn embedding by seen class proportions
	SYNC [3]	Learn base classifiers→unseen class classifiers

Per-class Top-1 Accuracy:

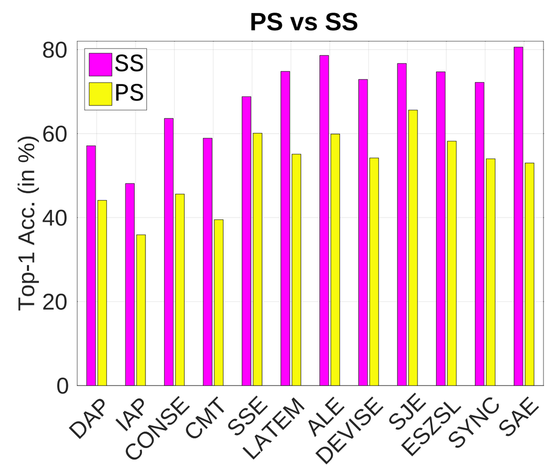
$$acc_{\mathcal{Y}} = \frac{1}{\lVert\mathcal{Y}\rVert} \sum_{c=1}^{\lVert\mathcal{Y}\rVert} \frac{\#\text{ correct in c}}{\#\text{ in c}}$$

Harmonic Mean:

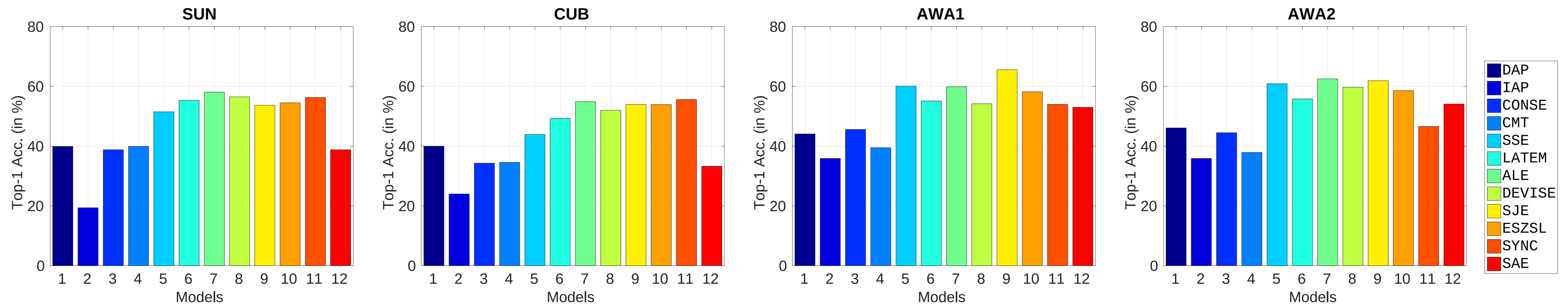
$$H = \frac{2 * acc_{\mathcal{Y}^{tr}} * acc_{\mathcal{Y}^{ts}}}{acc_{\mathcal{Y}^{tr}} + acc_{\mathcal{Y}^{ts}}}$$

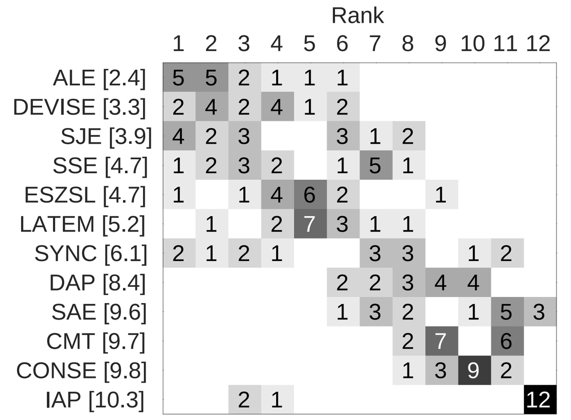
Propose a unified evaluation protocol and data splits
Compare and analyze a significant number of the state-of-the-art methods in depth
Zero-shot setting & more realistic generalized zero-shot setting

Zero-Shot Learning Setting

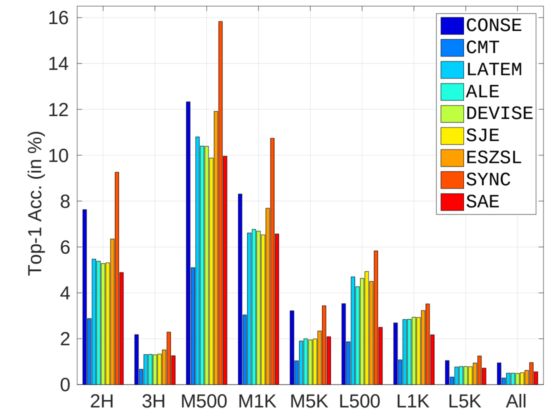


- Proposed Split (PS) vs Standard Split (SS) on AWA1
- 6 out of 10 test classes of SS appear in ImageNet 1K
- This violates zero-shot setting: Feature learning = part of training
- PS test sets are not a part of ImageNet 1K



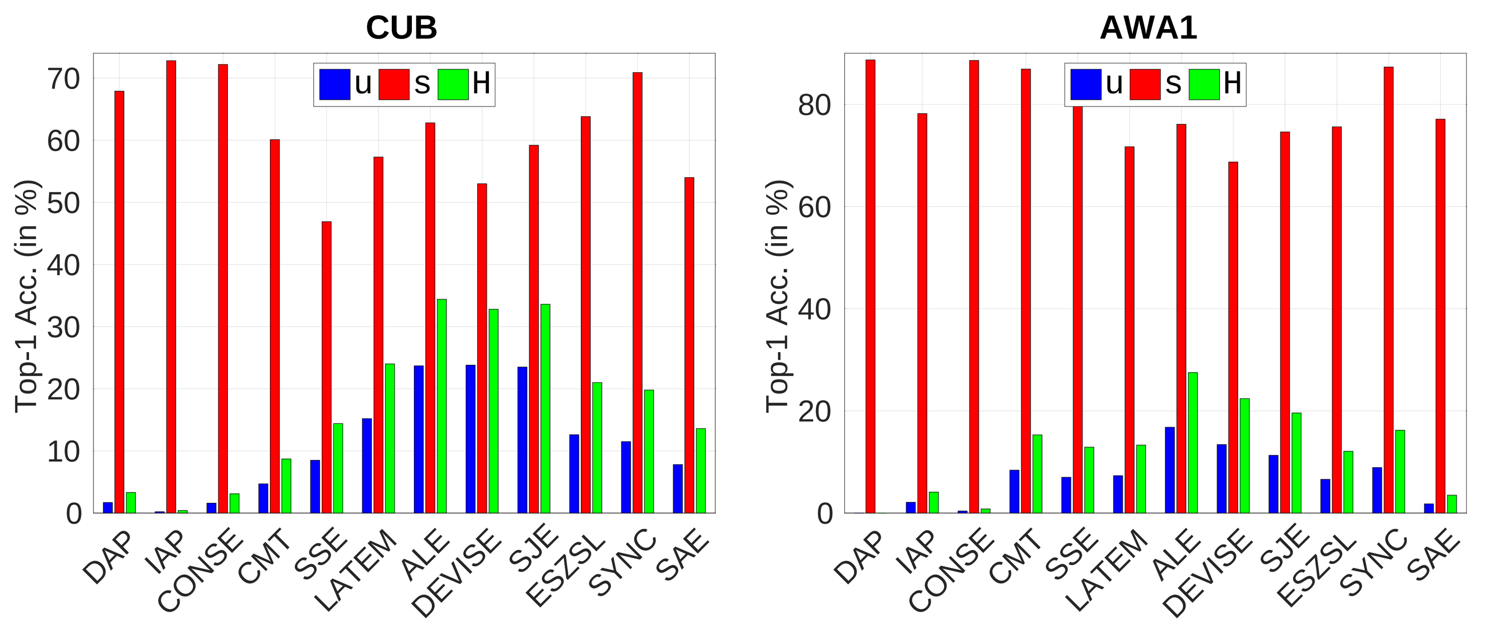


- Top-1 accuracy of PS on SUN, CUB, AWA1 and AWA2 (above)
- Methods ranking of PS on all 5 datasets (left)
  Element (i, j) indicates number of times model i ranks at jth
- Compatibility learning methods perform better

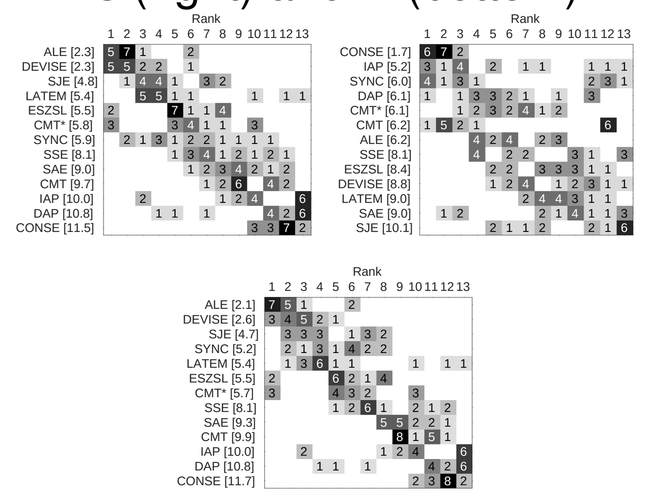


- Top-1 accuracy on ImageNet with different test class splits
- Hybrid models perform the best for highly populated classes
- All methods perform similarly for sparsely populated classes
- Results on least populated class lower than most populated class

Generalized Zero-Shot Learning Setting

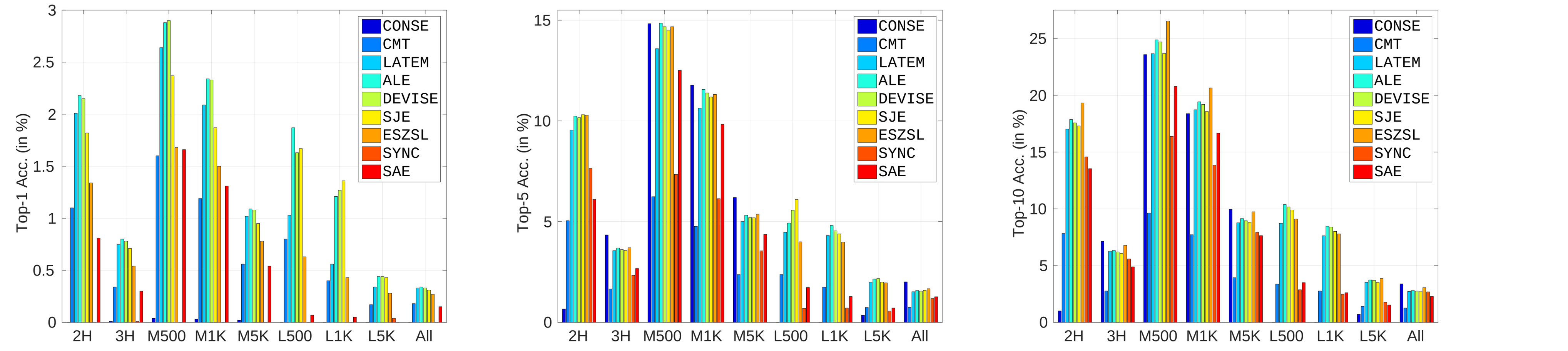


Ranking methods wrt u (left), s (right) and H (bottom)



- Methods with very high seen class accuracy have low unseen class accuracy
- Harmonic mean is a good measure that balances seen and unseen class accuracy

Top-K accuracy on ImageNet



[1] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid. Label-embedding for image classification. *TPAMI*, 2016.
[2] Z. Akata, S. Reed, D. Walter, H. Lee, and B. Schiele. Evaluation of output embeddings for fine-grained image classification. In *CVPR*, 2015.
[3] S. Changpinyo, W.-L. Chao, B. Gong, and F. Sha. Synthesized classifiers for zero-shot learning. In *CVPR*, 2016.
[4] A. Frome, G. S. Corrado, J. Shlens, S. Bengio, J. Dean, M. A. Ranzato, and T. Mikolov. Devise: A deep visual-semantic embedding model. In *NIPS*, 2013.
[5] E. Kodirov, T. Xiang, and S. Gong. Semantic autoencoder for zero-shot learning. In *CVPR*, 2017.
[6] C. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot visual object categorization. In *TPAMI*, 2013.
[7] M. Norouzi, T. Mikolov, S. Bengio, Y. Singer, J. Shlens, A. Frome, G. Corrado, and J. Dean. Zero-shot learning by convex combination of semantic embeddings. In *ICLR*, 2014.
[8] B. Romera-Paredes and P. H. Torr. An embarrassingly simple approach to zero-shot learning. *ICML*, 2015.
[9] R. Socher, M. Ganjoo, C. D. Manning, and A. Ng. Zero-shot learning through cross-modal transfer. In *NIPS*. 2013.
[10] Y. Xian, Z. Akata, G. Sharma, Q. Nguyen, M. Hein, and B. Schiele. Latent embeddings for zero-shot classification. In *CVPR*, 2016.
[11] Y. Xian, H. C. Lampert, B. Schiele, and Z. Akata. Zero-shot learning - a comprehensive evaluation of the good, the bad and the ugly. *arXiv preprint arXiv:1707.00600*, 2017.
[12] Z. Zhang and V. Saligrama. Zero-shot learning via semantic similarity embedding. In *ICCV*, 2015.