Semi-supervised Vocabulary-informed Learning
(Supplementary Material)

Yanwei Fu and Leonid Sigal
Disney Research
y.fu@qmul.ac.uk, lsigal@disneyresearch.com

1. Qualitative open set recognition results

We show some qualitative results of AwA (in Table 2) and ImageNet2012/2010 dataset (in Table 1) respectively. For each image, top-10 predicted labels produced by SVR-Map and SS-Voc:W are compared. This visually explains why our framework is better than SVR-Map. For example, the *siamese cat* in Table 2 has similar visual appearance, and highly similar semantic word vectors, in our dictionary, with *persian_cat* and *chihuahua dog*. SVR-Map is hence confused by these three animals. In contrast, our framework optimizes the mapping weight W with the nearest neighbor open vocabulary. Specifically, it will maximize W by enforcing the training instances of *siamese cat* class are closer to the prototypes of *siamese cat* but farther away from *persian cat*. Another interesting example comes from the *tripod* class in Table 1. The image is visually similar to *theater curtain* or *window screen*. Our SS-Voc:W maximizes the projection W to better separate the *tripod* class from the other classes using pairwise vocabulary max-margin terms.

2. Insensitivity to parameter settings

Performance of the proposed method as a function of parameters (μ, λ and α) is illustrated in Figure 3. We note that across wide variety of settings the performance (illustrated along the vertical axis) remains largely unchanged with only minor perturbations. This shows that our approach is very stable with respect to the parameter settings.

3. Larger version of figures

We also supply larger versions of Figures 1 and 3, from the main paper, for better readability.

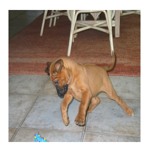
ImageNet	SVR-Map	SS-Voc:W
	passenger car, vehicle plane, airplane, tow truck **airliner**, car, aircraft truck, freight car	**airliner**, plane, airplane, passenger car, aircraft, bomber, vehicle, jet helicopter, lnb
	flat_coated_retriever **rhodesian ridgeback** bedlington_terrier weimaraner, log_jul irish_water_spaniel naya_rivera, libyan_officials shetland_sheepdog german_wirehaired_pointer	**rhodesian ridgeback** bedlington_terrier vizsla flat_coated_retriever log_jul, aloys, irish_water_spaniel merchant_seamen weimaraner, naya_rivera
	theater curtain, curtain front, wall clock, table lamp, **tripod**, car mirror, frac, provide_easy_access cowards	**tripod**, window screen window shade, table lamp curtain, car mirror window, arema motor_vehicle_manufacturers nominal

Table 2. Qualitative results of ImageNet 2012/2010 dataset with Top-10 prediction of each method. Ground truth names are in bold.

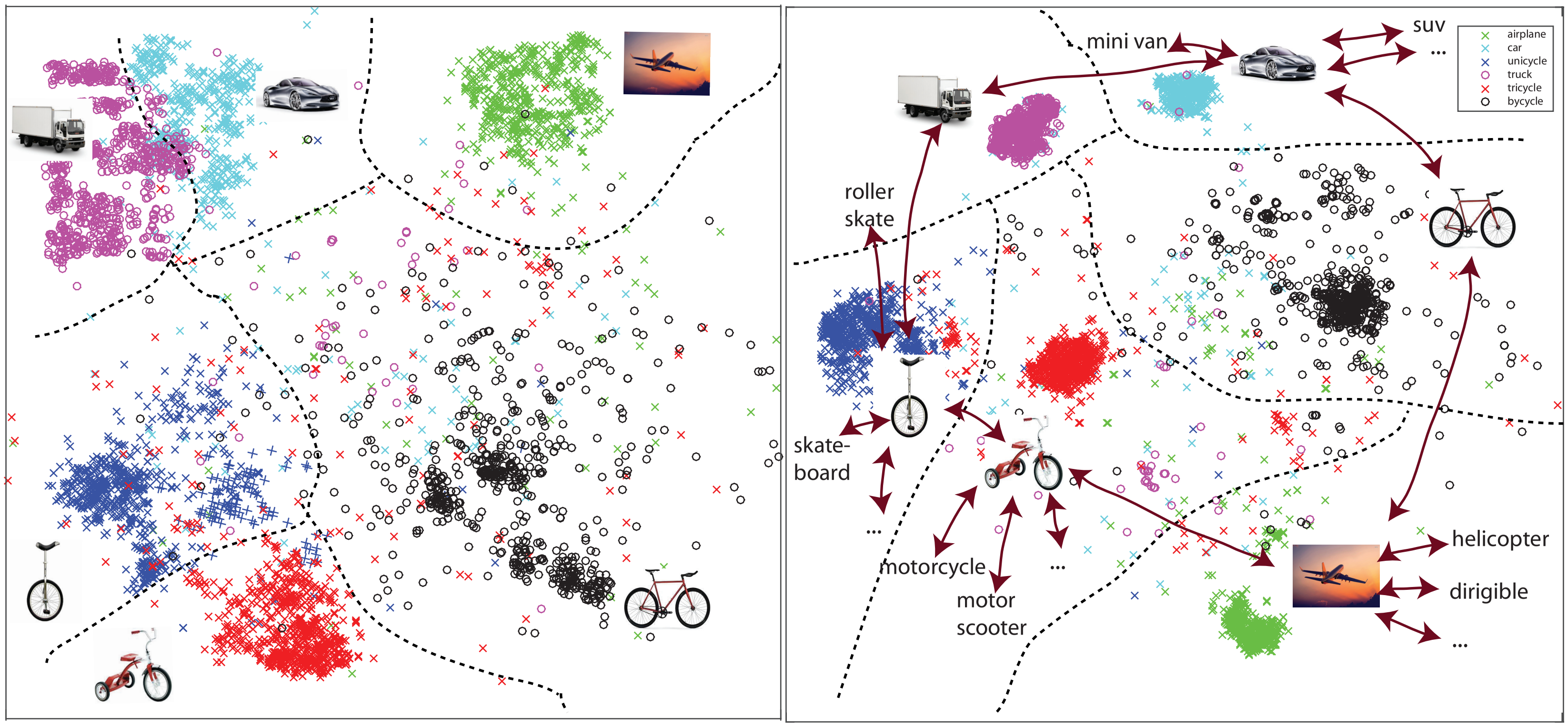


Figure 1. **Illustration of the semantic embeddings** learned (left) using support vector regression (**SVR**) and (right) using the proposed semi-supervised vocabulary-informed (**SS-Voc**) approach. In both cases, t-SNE visualization is used to illustrate samples from 4 source/auxiliary classes (denoted by $\times$) and 2 target/zero-shot classed (denoted by $\circ$) from the ImageNet dataset. Decision boundaries, illustrated by dashed lines, are drawn by hand for visualization. Note, that (i) large margin constraints in SS-Voc, both among the source/target classes and the external vocabulary atoms (denoted by arrows and words), and (ii) fine-tuning of the semantic word space, lead to a better embedding with more compact and separated classes (*e.g.*, see *truck* and *car* or *unicycle* and *tricycle*).

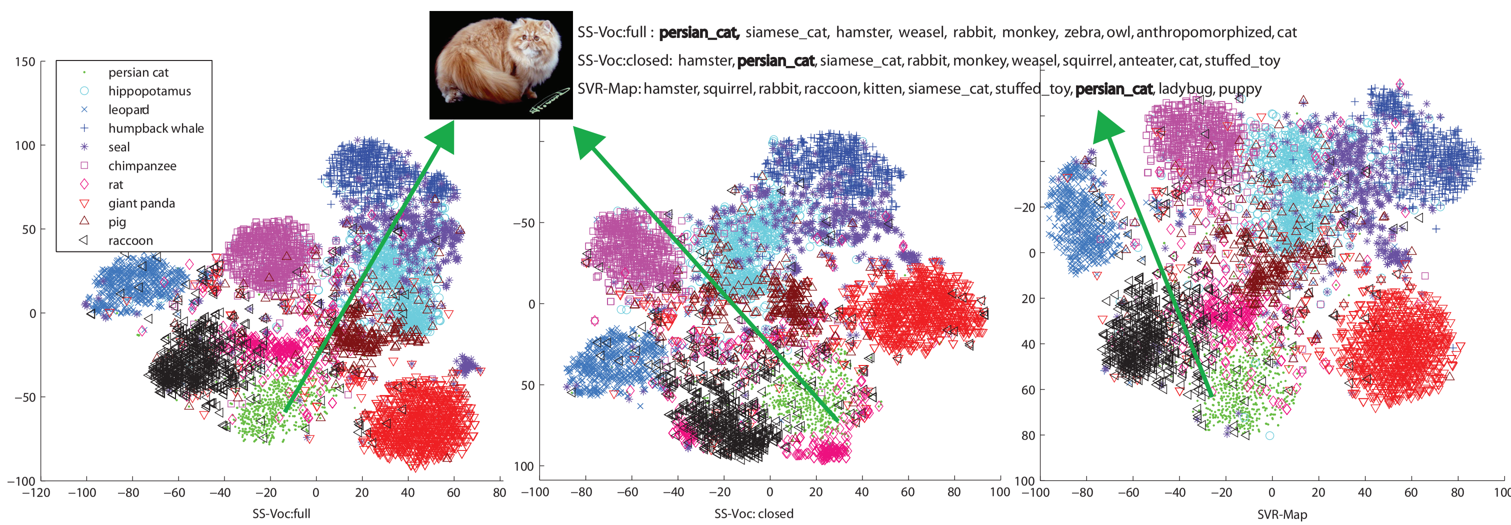


Figure 2. t-SNE visualization of AwA 10 target classes

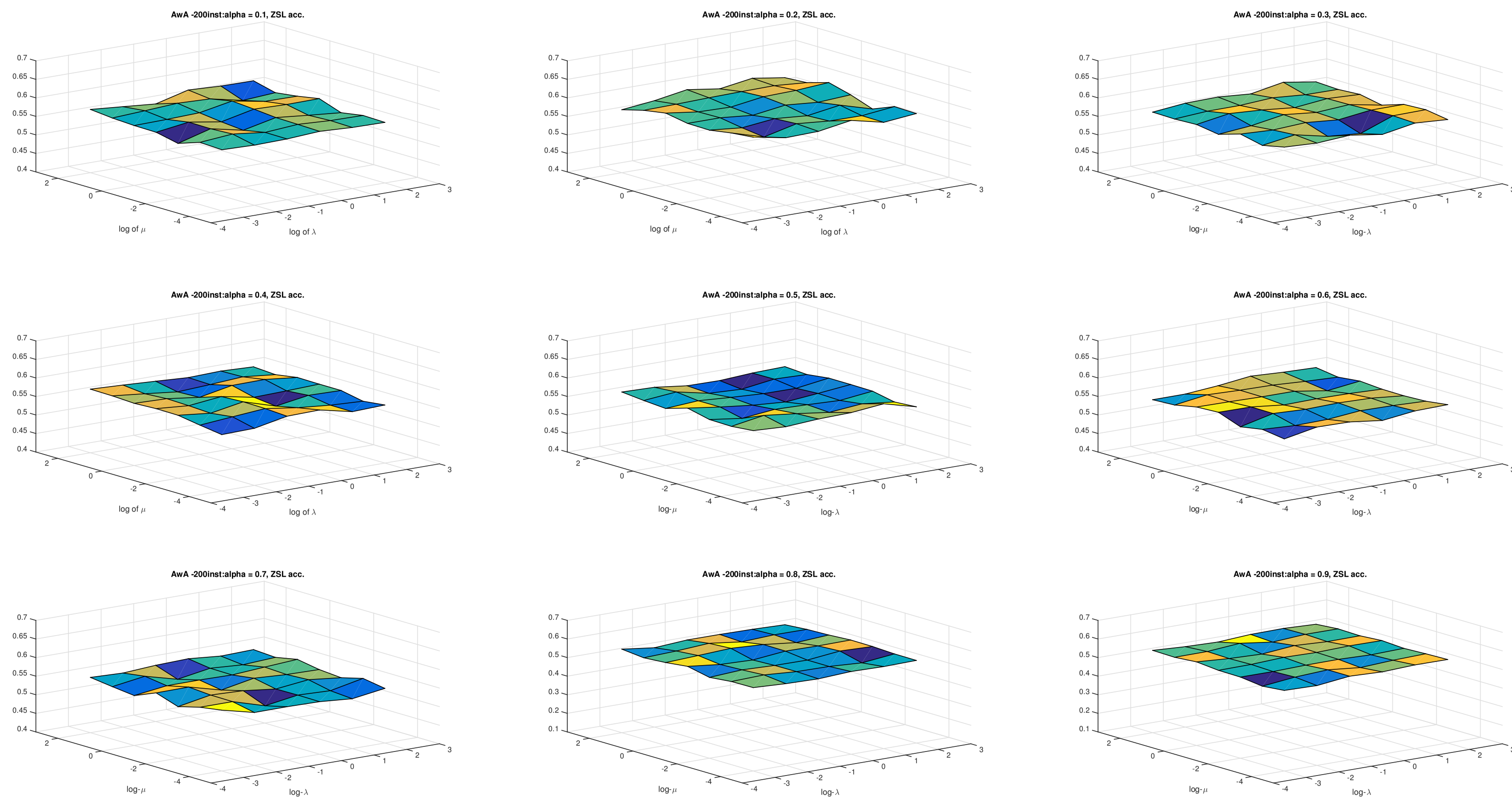


Figure 3. Zero-shot performance (vertical-axis) on 10 AwA classes as a function of parameter values: λ (x-axis), μ (y-axis) and α (different plots). Each experiment is repeated 10 times and results are averaged. Note that accuracy is highly invariant / insensitive to the exact parameter settings.

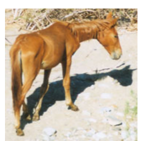
AwA	SVR-Map	SS-Voc: W
	elephant, giraffe lioness, monkey crocodile, hyena pachyderm, tortoise gorilla, asian elephant	**horse**, elephant hoofs, hooves bull, goat lioness, donkey herd, horseman
	chihuahua dog, persian cat raccoon, squirrel **siamese cat,** shih tzu kitten, budgies, cockatiel jack russell terrier	**siamese cat**, persian cat chihuahua dog, cat rabbit, kitten jack russell terrier hamster, shih tzu, chow chow

Table 1. Qualitative results of AwA dataset with Top-10 prediction of each method. Ground truth names are in bold.