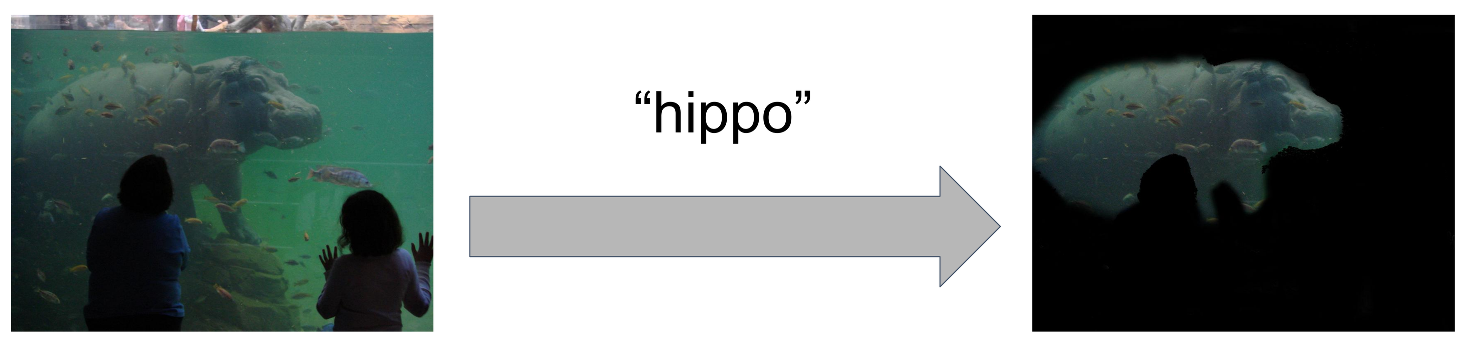
LEARNING SEMANTIC-GUIDED VISUAL ATTENTION FOR FEW-SHOT IMAGE CLASSIFICATION

Wen-Hsuan Chu, Yu-Chiang Frank Wang

Department of Electrical Engineering, National Taiwan University, Taiwan

MOTIVATION

- Semantic information in labels contain information that helps humans learn new concept. Can machines also exploit this information to learn more efficiently especially when training data is limited?



Object 5? Where? Ah ha!

CONTRIBUTIONS

- Our proposed model incorporates semantic information in the labels to guide the attention models in an **unsupervised** fashion.
- We use the attended images and its complement as a form of **data augmentation** to help combat the scarcity of training images in few-shot settings.
- Our method performs favorably compared to baselines on multiple datasets, AwA2 and CIFAR-100, and show consistent improves as we increase the amount of training images.

PROPOSED METHOD

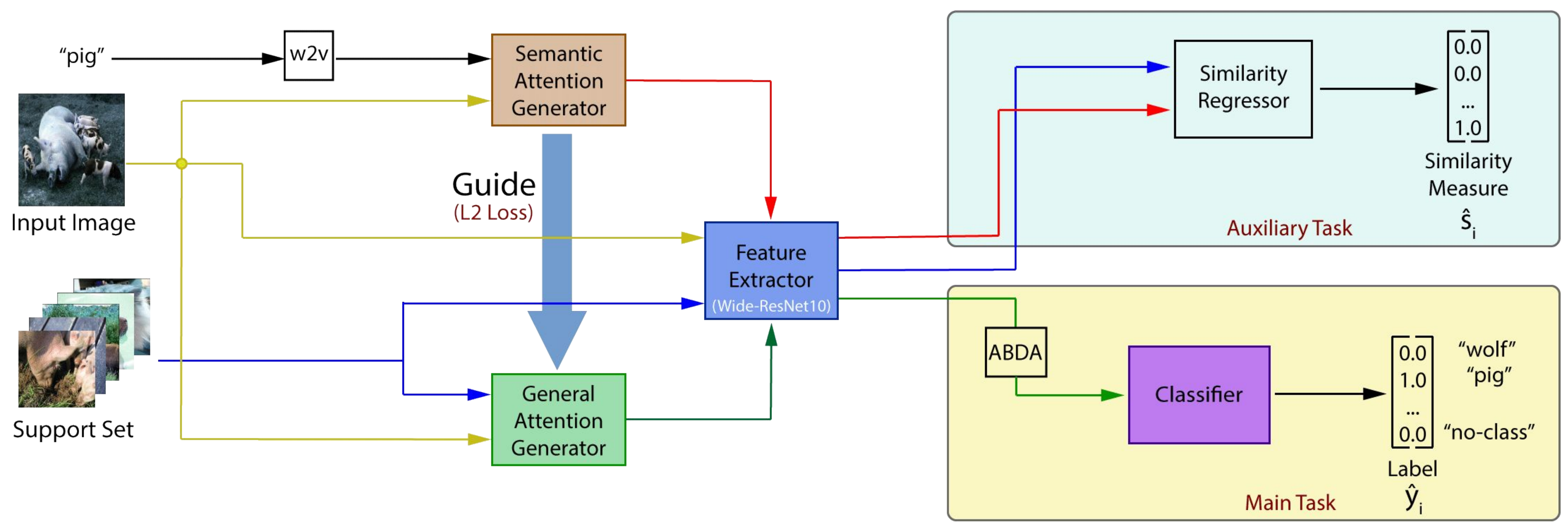


Fig. Few-shot learning with **semantic-guided attention**.

TRAINING

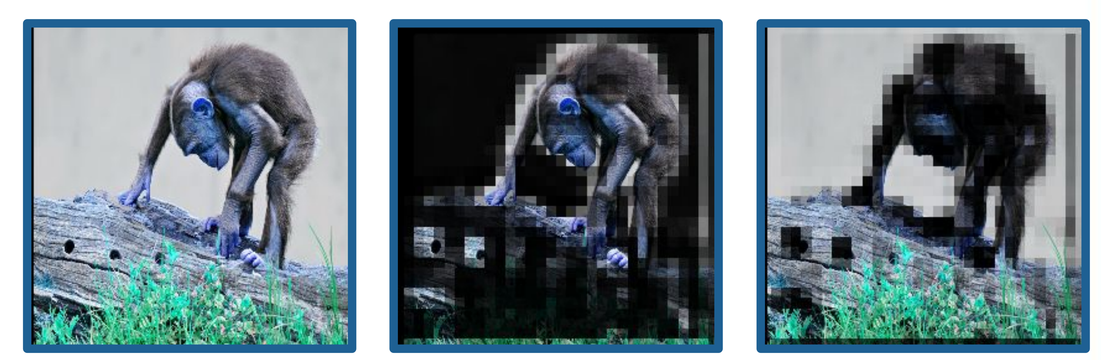
- Auxiliary Task measures the similarity between the **general attended support set** and the **semantic attended input image**
- Main Task outputs the label based on the **general attended input image**
- We let the two attention map outputs be as similar as possible via a L2 loss

TESTING

- Label is predicted based on the **general attended input image** as no labels are available

Attention-Based Data Augmentation (ABDA)

- Use the attention images and its complements as a form of data augmentation by creating a "no-class" label



Original Image — Positive Sample — Negative Sample

Gorilla — Gorilla — No-Class

EXPERIMENTS

Datasets

- ***AwA2***: 40 base training classes, randomly sample K images from the 10 testing classes as "few-shot" training samples
- ***CIFAR-100***: 80 base training classes, 20 "few-shot" training classes

Experiment Settings

- (a) ***Closed-world***: evaluate on new "few-shot" classes only
- (b) ***Open-world***: evaluate on all training classes

Ablation Studies

- 1. ***Baseline***: CNN model only
- 2. ***Center***: CNN model with center loss as regularization
- 3. ***Proposed***: Our method

Table 1: Results on CIFAR-100 for the closed-world setting.

	k=1	k=2	k=5
Baseline	22.25%	31.20%	42.00%
Center	24.05%	36.35%	44.25%
Proposed	25.20%	38.40%	43.80%

Table 2: Results on CIFAR-100 for the open-world setting.

	k=1	k=2	k=5
Baseline	10.67%	17.35%	29.61%
Center	9.82%	18.10%	31.65%
Proposed	11.36%	21.68%	34.88%

Table 3: Results on AwA2 for the closed-world setting.

	k=1	k=2	k=5
Baseline	31.78%	33.81%	42.53%
Center	37.34%	37.92%	48.80%
Proposed	40.97%	42.80%	55.06%