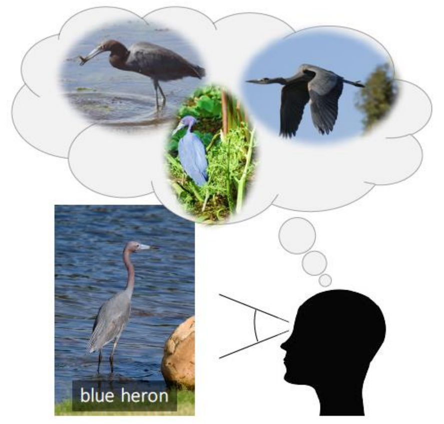
Semantics-guided Data Hallucination for Few-shot Visual Classification

Chia-Ching Lin[1,2] Yu-Chiang Frank Wang[1] Chin-Laung Lei[1] Kuan-Ta Chen[2]

[1]Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
[2]Institute of Information Science, Academia Sinica, Taipei, Taiwan

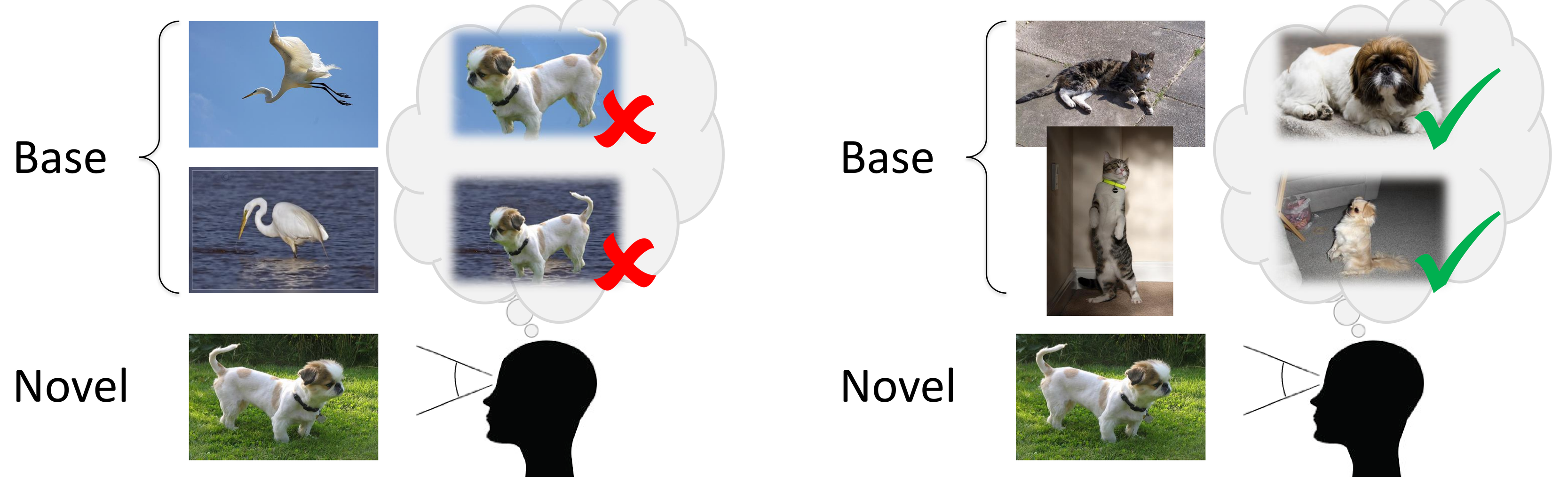
INTRODUCTION

- Few-shot learning (FSL): only few samples would be available for selected object categories during learning
  - **Base** classes: each has a sufficient amount of training samples
  - **Novel** classes: each has a limited amount of training samples
- **Data hallucination**: generate additional training samples for novel classes based on intra-class variation learned from base classes



(CVPR18)

- Idea: base classes with similar **semantic concepts** to the underlying novel class should be considered for reasonable hallucination



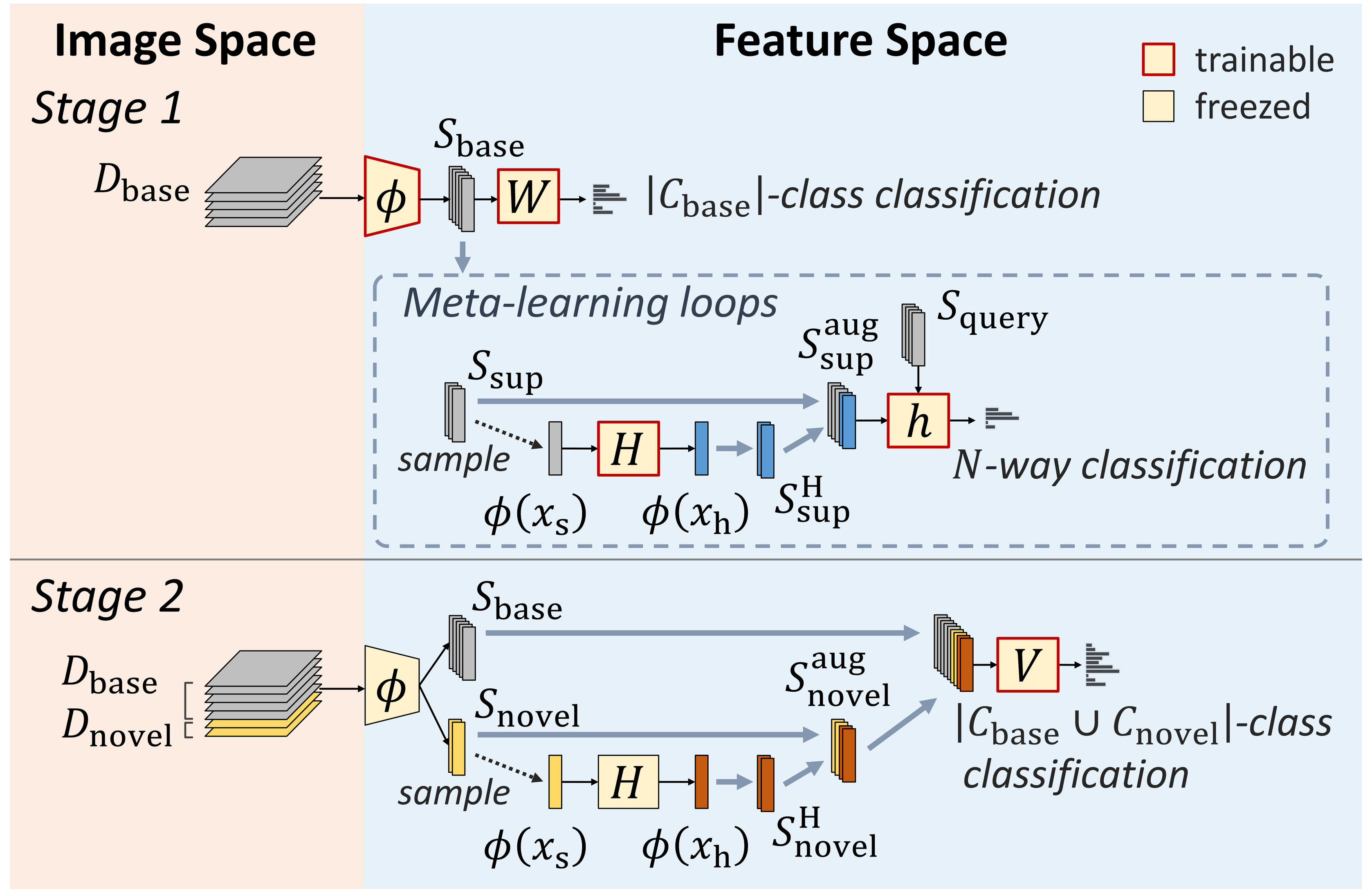
- Does more reasonable hallucination lead to better FSL performance? If so, how to achieve that?

NOTATIONS

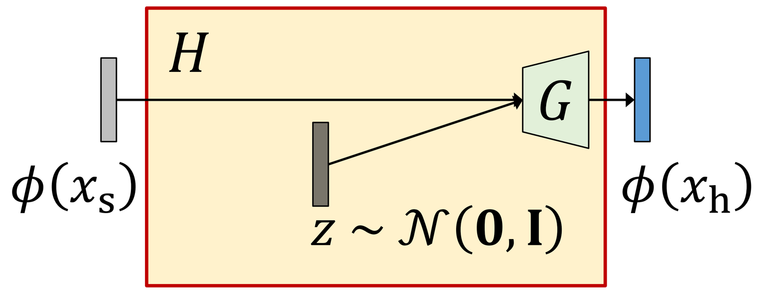
- D_{base}: dataset of base classes C_{base}
- D_{novel}: dataset of novel classes C_{novel}
- Both datasets consist of tuples $\{(x_i, y_i, R_i)\}$
  - x_i: the i-th image
  - y_i: the one-hot label vector
  - R_i: the **semantic information** associated with y_i
    - CIFAR-100: a word embedding vector of the label name
    - Animals with Attributes: an attribute vector

PROPOSED MODEL

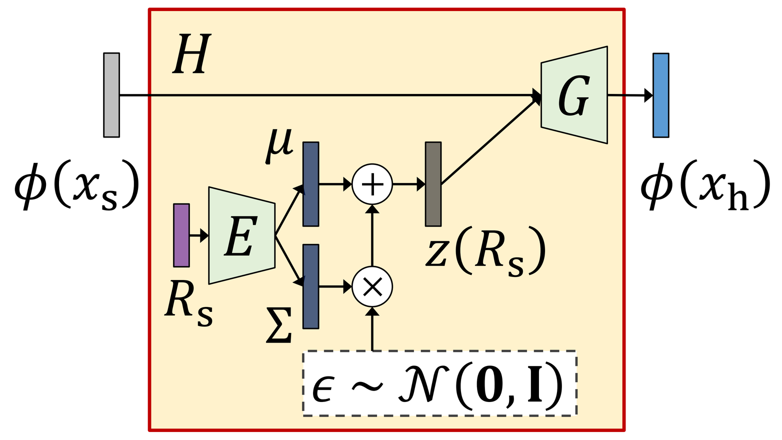
- Training
  - *Stage 1* (representation learning on D_{base}): train the feature extractor ϕ and the **hallucinator** H by meta-learning
  - *Stage 2* (few-shot learning on $D_{\text{base}} \cup D_{\text{novel}}$): use the trained H to augment additional training samples for C_{novel} and use the augmented dataset to train the final classifier V



- H proposed in (CVPR18) models intra-class variation by noise vectors sampled from a fixed distribution: $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- Our H incorporates semantic information R_s to produce z as if they are sampled from a **semantics-dependent** distribution



(CVPR18)



(Ours)

- Inference: Use ϕ and V to predict the label of an unseen test image x from $C_{\text{base}} \cup C_{\text{novel}}$

EXPERIMENTS

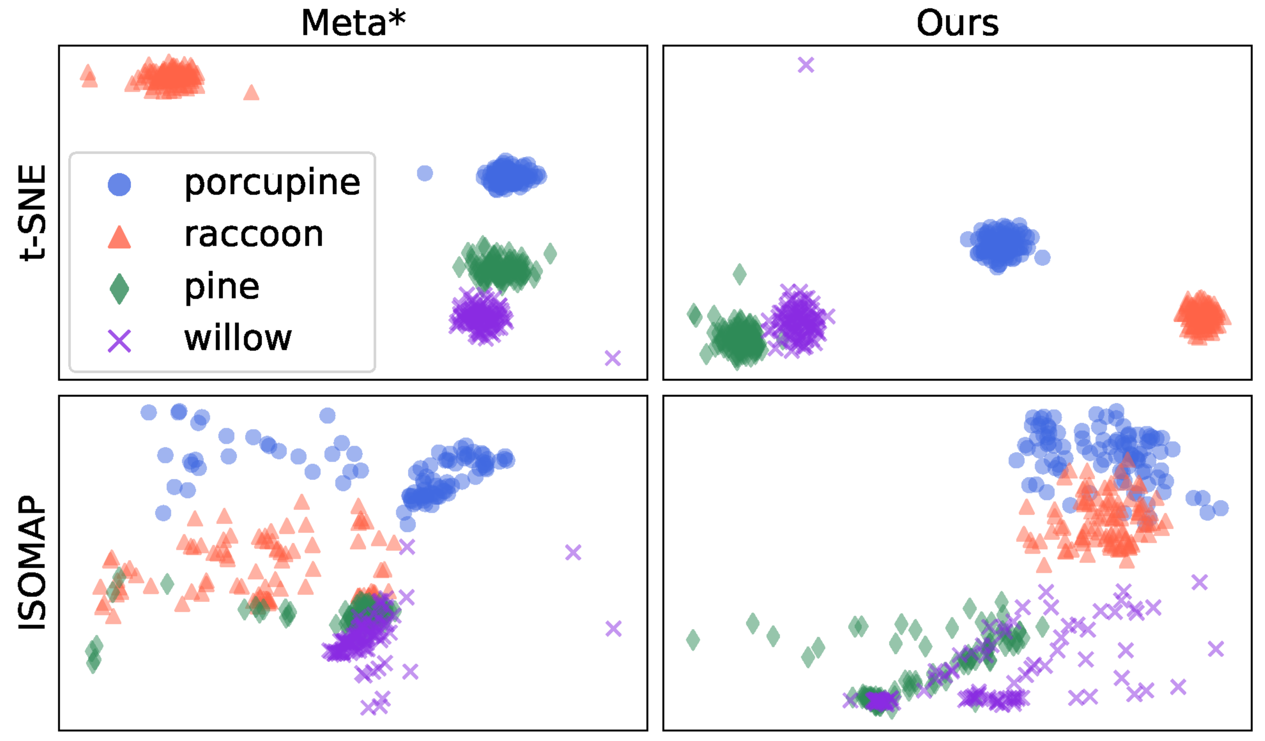
- Top-5 accuracy of V
  - CIFAR-100: $|C_{\text{base}}^{\text{fin}}| = 30$, $|C_{\text{novel}}^{\text{fin}}| = 20$

	$C_{\text{novel}}^{\text{fin}}$			$C_{\text{base}}^{\text{fin}} \cup C_{\text{novel}}^{\text{fin}}$		
	$n=1$	$n=2$	$n=5$	$n=1$	$n=2$	$n=5$
Baseline	28.96	46.20	60.85	67.36	73.59	80.35
Meta (CVPR18)	47.65	59.65	70.82	70.88	77.07	81.86
Meta* (CVPR18)	49.95	58.64	71.15	74.36	77.60	81.67
Analogy (ICCV17)	38.51	49.76	65.27	71.65	75.83	81.34
Superclass	46.76	57.01	67.86	74.65	78.44	82.30
Ours	**55.84**	**66.28**	**74.14**	**76.68**	**80.18**	**82.95**

  - Animals with Attributes: $|C_{\text{base}}^{\text{fin}}| = 15$, $|C_{\text{novel}}^{\text{fin}}| = 10$

	$C_{\text{novel}}^{\text{fin}}$			$C_{\text{base}}^{\text{fin}} \cup C_{\text{novel}}^{\text{fin}}$		
	$n=1$	$n=2$	$n=5$	$n=1$	$n=2$	$n=5$
Baseline	42.03	56.93	72.78	71.03	77.70	84.85
Meta (CVPR18)	59.04	67.14	76.61	77.28	80.98	84.86
Meta* (CVPR18)	58.96	67.05	75.93	77.53	80.67	84.27
Ours	**63.59**	**70.55**	**78.20**	**79.78**	**82.50**	**85.40**

- 2D visualization of hallucinated features on selected classes of CIFAR-100



CONCLUSION

- We incorporate semantic information into the data hallucination process to generate additional training data that exhibit semantics-oriented modes of variation for improved FSL performances