Low-Shot Learning from Imaginary Data

Yu-Xiong Wang[1,2] Ross Girshick[1] Martial Hebert[2] Bharath Hariharan[1,3]
[1]Facebook AI Research (FAIR) [2]Carnegie Mellon University [3]Cornell University

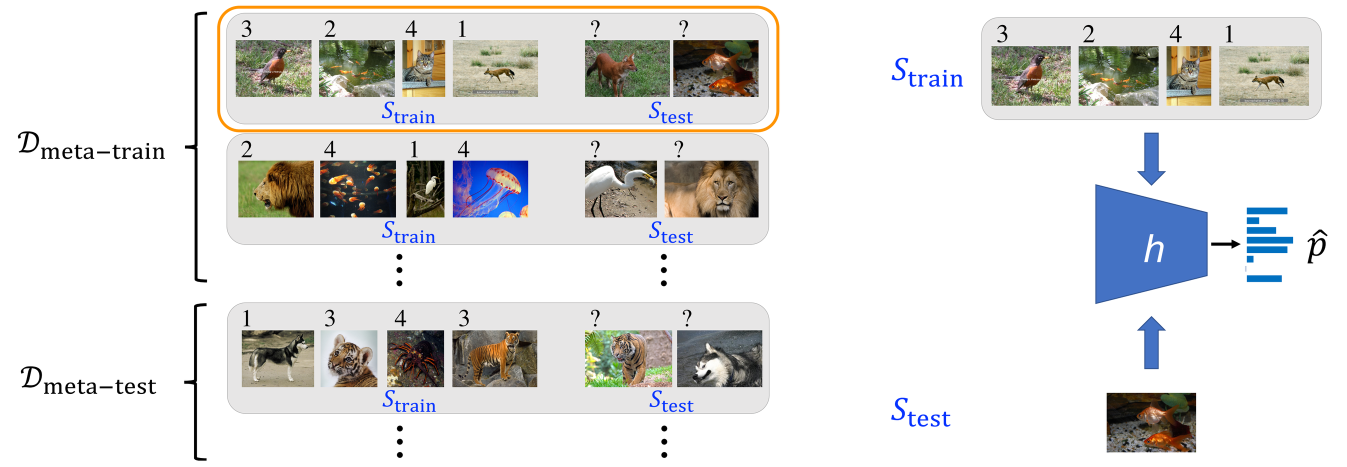
CVPR 2018 SALT LAKE CITY • JUNE 18-22

Motivation

- **Problem statement:** low/few-shot learning
- **Key insight:** leverage structure of the visual world
  - Modes of variation are shared across classes
  - Humans can visualize a novel object in other poses or surroundings
- **Our approach:** an end-to-end model that "hallucinates" novel training samples
  - **Hallucination criterion:** produce examples useful for classification
  - **Training hallucinator with meta-learning:** *jointly* optimize a meta-learner [1] with a hallucinator

Meta-Learning

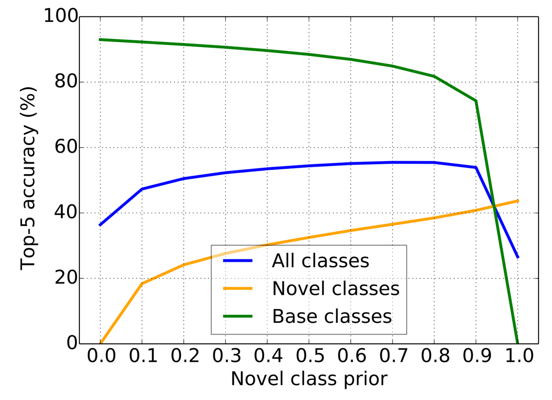
- **Setup:** learn from a meta-set—a collection of low-shot classification tasks
- **Meta-training:** train a classification algorithm h on small S_{train} that achieves high accuracy on S_{test} for different tasks on $\mathcal{D}_{\text{meta-train}}$
- **Meta-testing:** use h to solve novel classification tasks on $\mathcal{D}_{\text{meta-test}}$



- **Flagship techniques:** prototypical networks (MLP embeddings, prototype/class mean classifiers), matching networks (LSTMs contextual embeddings, nearest neighbor classifiers)
- **Our prototype matching networks:** LSTMs contextual embeddings, prototype classifiers

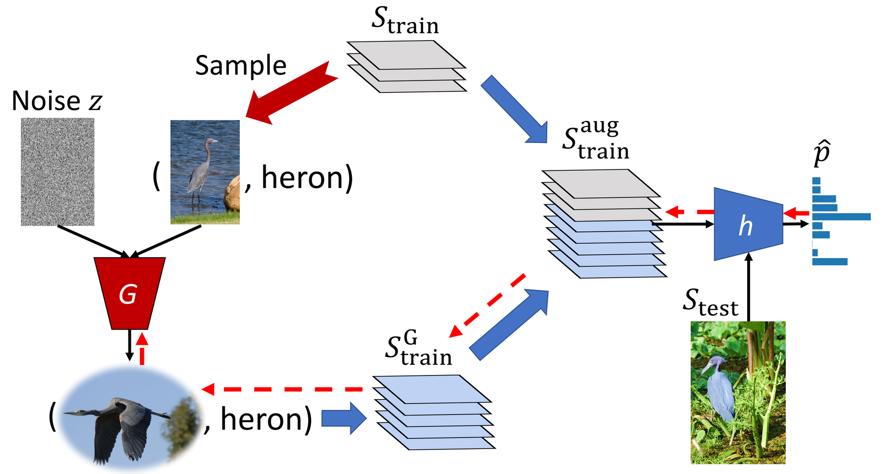
Experimental Protocol

- **Tradeoffs between base and novel classes in joint evaluation:** a novel class prior μ
- **A new evaluation:** top-5 accuracy
  - in only-novel label space
  - in only-base label space
  - in joint label space *without* and *with* a cross-validated μ



Meta-Learning with Learned Hallucination

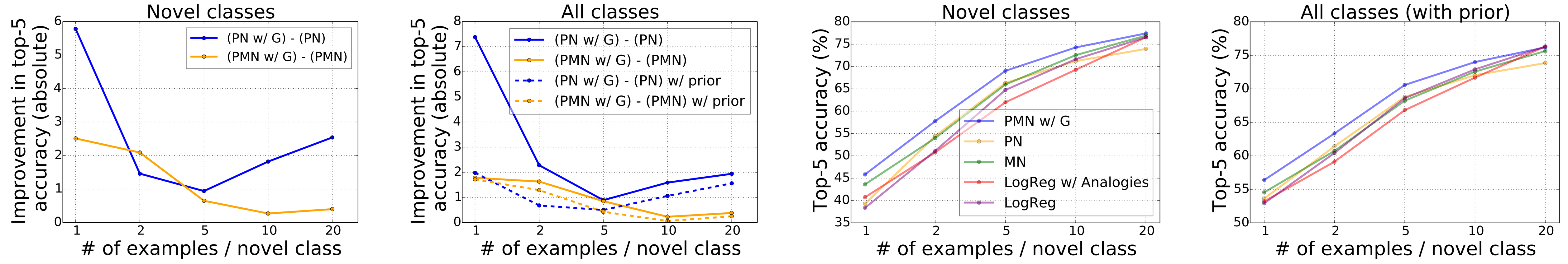
- **A general framework:** agnostic to meta-learning algorithms
- **Hallucinator:** parametric function $G(x, z; \mathbf{w}_G)$
- **Meta-training the hallucinator**
  - **Augmented training set** $S_{\text{train}}^{\text{aug}}$: $S_{\text{train}} \cup S_{\text{train}}^{G}$
  - **End-to-end training:** G along with h
- **Benefits of end-to-end training**
  - Directly produces hallucinations useful for class distinctions
  - Makes allowances for any errors in the hallucination
  - No extra annotation & no heuristics



(Hallucination is performed in feature space; images shown for illustration.)

ImageNet Low-Shot Classification

- **Benchmark:** $C^{\text{cv}} = C^1_{\text{base}}$ (193 classes) $\cup\, C^1_{\text{novel}}$ (300 classes) | $C^{\text{test}} = C^2_{\text{base}}$ (196 classes) $\cup\, C^2_{\text{novel}}$ (311 classes)
- **State-of-the-art meta-learners:** prototypical networks (PN) | matching networks (MN) | *our* prototype matching networks (PMN)
- **Hallucinator** G: three layer MLP with ReLU
- **Pre-trained features:** ResNet-10 and ResNet-50
- **Impact of hallucination**
  - Significant improvements on novel classes
  - General with different meta-learners
  - Comparably good on base classes, 92% top-5 accuracy
- **Comparison to prior work**
  - **Baselines:** logistic regression | logistic regression with analogies hallucination
  - **Resilience to mis-calibration:** gains from hallucination decrease when μ is cross-validated



- **Consistent improvements using deeper ResNet-50 features**

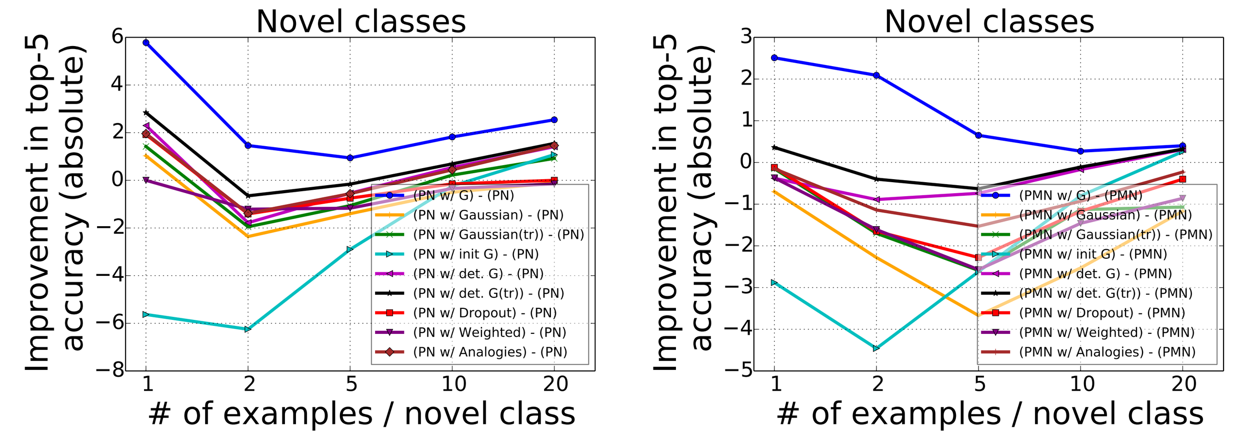
ResNet-50	n=1	2	5	10	20
PMN w/ G*	**54.7**	**66.8**	**77.4**	**81.4**	**83.8**
PMN*	53.3	*65.2*	*75.9*	80.1	82.6
PN w/ G*	*53.9*	65.2	75.7	*80.2*	*82.8*
PN	49.6	64.0	74.4	78.1	80.0
MN	53.5	63.5	72.7	77.4	81.2

- **Competitive performance with only hallucinated examples**

ResNet-10	n=1	2	5	10	20
PN w/ G*	**45.0**	**55.9**	**67.3**	**73.0**	**76.5**
PN only G*	*42.4*	*54.7*	65.9	70.8	73.5
PN	39.3	54.4	*66.3*	*71.2*	*73.9*

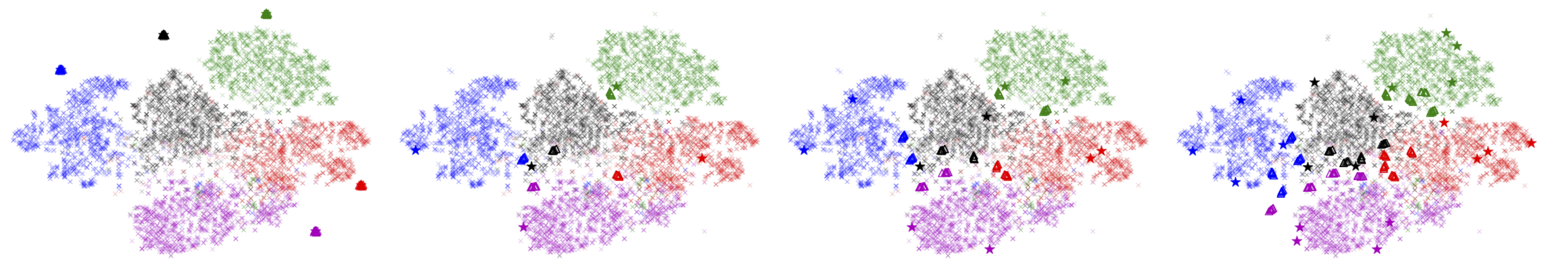
Unpacking the Performance Gain

- **Sophisticated hallucination architectures** are necessary
- **Meta-learning the hallucinator** is necessary
- **Diverse samples** are produced



Visualizing the Learned Hallucinations

- **t-SNE visualizations:** on novel classes for prototypical networks
- **Our hallucinator vs. baseline Gaussian hallucinator:** match the class distributions more closely & with different seed examples capture different parts of the space
- **Clustering around the class boundaries:** perhaps a consequence of discriminative training of the hallucinator



Gaussian baseline | G with 1 seed | 2 seeds | 4 seeds

×: real examples; ⋆: seeds; △: hallucinations

Conclusion and Future Work

- Low-shot learning that uses a hallucinator to generate additional examples and trains the hallucinator *end-to-end* with meta-learning
- Significant gains *irrespective of* meta-learning approaches
- **Future work:** pin down exactly the effect of the hallucinated examples & increase the diversity of the hallucinations

References

1. O. Vinyals, *et al.* Matching networks for one shot learning. In NIPS, 2016.
2. B. Hariharan and R. Girshick. Low-shot visual recognition by shrinking and hallucinating features. In ICCV, 2017.