

MOTIVATION

- Intrinsic *long-tailed* distribution for recognition tasks in the wild
- Head-to-tail meta-knowledge transfer
- Meta-level network: Operate on the space of model parameters
- Model dynamics: Transformations from few-shot to many-shot models
- Progressive transfer
- A **single**, **chained** MetaModelNet for models of different sample sizes
- **Recursive** class splitting into head and tail
- An illustration: Learn both many-shot and few-shot living-room models, and train a **regressor** that maps between the two



REGRESSING k-Shot to Many-Shot Models

• Base learner and meta-learner

- Training set H_t of head classes: (x, y) data-label pairs for classes with more than t training examples
- Base learner $g(x; \theta)$: Feedforward function with parameters θ
- Optimal model parameters θ_* : Tuning g on H_t with a standard loss function
- Few k-shot model parameters θ : Tuning g on random subsets of H_t with k examples per class
- *MetaModelNet* $\mathcal{F}(\theta; w)$: Meta-network regressing θ to θ_* with parameters w

Loss function

$$\sum_{\theta \in k \operatorname{Shot}(H_t)} \left\{ ||\mathcal{F}(\theta; w) - \theta_*||^2 + \lambda \sum_{(x,y) \in H_t} \operatorname{loss}\left(g(x; \mathcal{F}(\theta; w)), y\right) \right\}$$

• Parameters θ and θ_*

- In principle: $\mathcal{F}(\cdot)$ applies to model parameters from multiple CNN layers
- Sharable across classes: Parameters from the classifier module (last fullyconnected layer) for a single class



[Wang & Hebert, ECCV, 2016]

LEARNING TO MODEL THE TAIL

Martial Hebert Deva Ramanan Yu-Xiong Wang Email: {yuxiongw, dramanan, hebert}@cs.cmu.edu

- models





CONCLUSIONS

- Long-tail recognition: Meta-knowledge transfer from data-rich head to data-poor tail classes
- Representation of model dynamics: How model evolves when gradually encountering more training examples
- Learning to learn: Regressing model parameters Progressive learning: Back-to-front residual learning **Curriculum learning:** Learning from classes in the head to the body and then to the tail



- Structure in *dual* model (parameter) space: Models $\theta \in \mathbb{R}^{2048}$ as points for ResNet • **Model evolution** with increasing sample sizes: Trajectories over the model space
- **Class-specific** data augmentation
- PCA: Approximately smooth, nonlinear warping of model space





ABLATION ANALYSIS

• Importance of **sample-size dependent** transformation and identity regular-

• Curriculum learning in the way of recursive head-tail class splitting further improves performance



• Fine-tuning the entire ResNet50



LARGE-SCALE COMPARISONS

- Scene classification on the Places dataset: Long-tailed version with 5–9,900 training
- Object classification on the ImageNet $\widehat{\otimes}_{60}$ dataset: Merging the 1,000 classes into the 200 higher-level classes
- **Base learner:** AlexNet trained from scratch
- Facilitate long-tail recognition with diverse visual concepts and distributions



UNDERSTANDING MODEL DYNAMICS



• t-SNE: Similar semantic classes tend to be close and transform in similar ways

