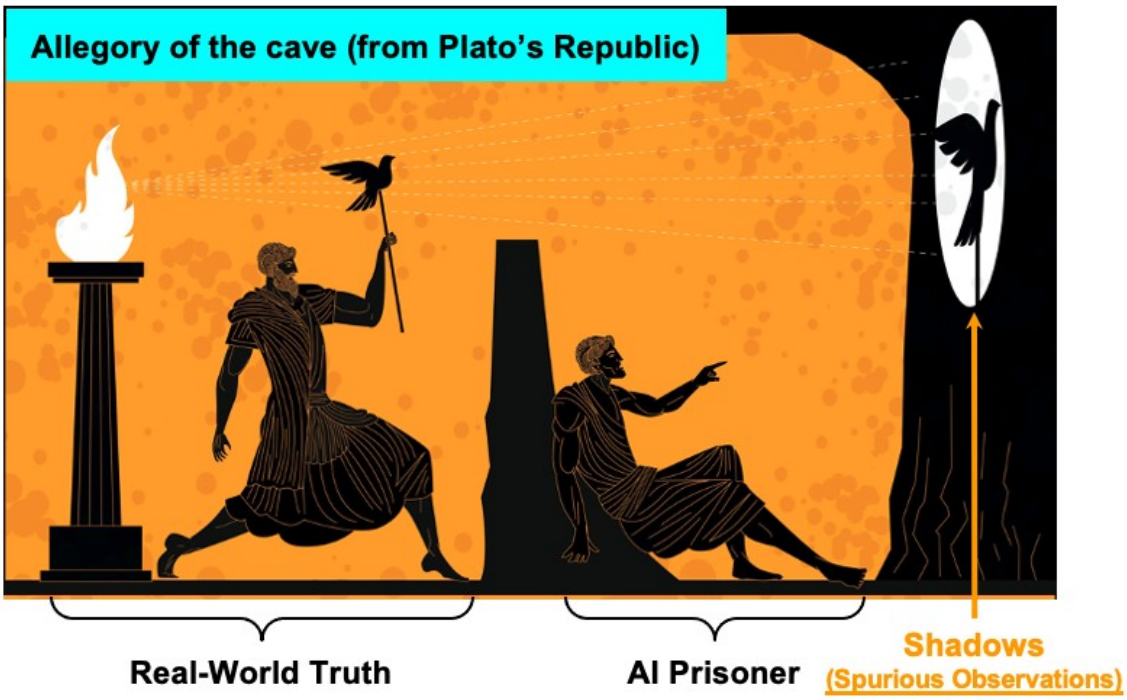
Research Statement

SUN Qianru
School of Computing and Information Systems,
Singapore Management University
Tel: (65) 6828-1360 Email: qianrusun@smu.edu.sg
01/01/2021

Background



My research interest is in computer vision and machine learning. More specifically, I am looking at how we can empower AI with causal reasoning and meta-learning abilities, so that it can grasp the true correlations among visual objects and understand the visual world.

The above picture is from the Plato's Allegory of the Cave. It described a group of people imprisoned in a cave, facing a blank wall, on which the shadows of outside objects were projected. Can the prisoners claim that they understand the world outside the cave by only watching those shadows? No, they can never. Hence, a question may arise - isn't today's data-driven AI models the same as the prisoners?

AI is good at watching tons of labelled images which are actually the shadows of our human lives and surroundings, and it beats us in various "shadow tricks'' such as memorization and imitation. Unfortunately, AI still fails in "trivial humanity'' such as common sense, sympathy, interpretability and generalization. To make AI truly understand the visual world and be able to tackle the real tasks robustly, we need to empower it to pursue the capabilities of causal reasoning and meta-learning (also called learning to learn). It is our hope that we will enable the AI models to

do efficient predictions out of large-scale supervised data but simply based on weakly supervised, few-shot or incremental data.

Research Areas

1. Meta-Learning is towards highly adaptative AI models!

What we need in meta-learning is a principled approach to generalizable knowledge or information among similar tasks, i.e., a way of sharing information between relevant learning processes. This is particularly useful in challenging learning scenarios such as few-shot learning and continual learning.

1.1 Few-shot learning

We expect machine learning model can learn new concepts/classes from a handful of training examples, e.g., from 1 or 5 training images [Sun et al. CVPR 2019]. Humans tend to be highly effective in such context, often grasping the essential connection between new concepts and their own knowledge and experience, but this still remains challenging for machine learning models. For instance, on the CIFAR-100 dataset, a classification model trained in the fully supervised mode achieves around 80% accuracy for the 100-class setting, while the best-performing 1-shot model achieves only around 60% in average for the much simpler 5-class setting. Besides, in many real-world applications we are lacking large-scale training data, as e.g. in the medical image domains. It is thus desirable to improve machine models in order to handle few-shot settings.

Learning to transfer knowledge. The nature of few-shot learning with very scarce training data makes it difficult to train powerful machine learning models for new concepts. One solution is to leverage the transferrable pattern learned from existing large-scale tasks. Our method is thus called Meta-Transfer Learning (MTL). First, large-scale trained DNN weights offer a good initialization, enabling fast convergence of MTL with fewer tasks. Second, light-weight transfer operations, i.e., scaling and shifting DNN neurons, have less parameters to learn, reducing the chance of overfitting to few-shot data. In a nutshell, MTL is a learning method that helps DNN converge faster while reducing the probability to overfit when training on few labeled data only. On the miniImageNet 1-shot case, we achieved 62.1% and 63.4% accuracies respectively on ResNet-18 and ResNet-25 backbones [Sun et al. PAMI 2020].

Learning to learn with unlabeled data. When there is not enough labeled data or annotation is costly, our solution is semi-supervised learning [Li et al. NeurIPS 2019]. It leverages unlabeled data and specifically meta-learns how to cherry-pick and label such unsupervised data to further improve performance. In specific we train the Learning to self-train (LST) model through a large number of semi-supervised few-shot tasks. On each task, we train a few-shot model to predict pseudo labels for unlabeled data, and then iterate the self-training steps on labeled and pseudo-labeled data with each step followed by fine-tuning. We additionally learn a soft weighting network (SWN) to optimize the self-training weights of pseudo labels so that better ones can contribute more to gradient descent

optimization. On the miniImageNet 1-shot learning case, we achieved 70.1% accuracy on a ResNet-12 backbone, when adding 30 times of unlabeled data to training.

Learning to ensemble models. The nature of few-shot learning with very scarce training data makes it difficult to train models stably. The model uncertainty is thus high and often results in low performance. For tackling this issue, we proposed the solution of training an ensemble of models and use the combined prediction which should be more robust. However, it is not obvious how to obtain and combine an ensemble of base-learners given the fact that a very limited number of training samples are available. Rather than learning multiple individual base-learners, we propose to use the sequence of base-learners while training a single base-learner as the ensemble and also learn how to weigh them for best performance automatically [Liu et al. ECCV 2020]. Second, it is well known that the values of various hyperparameters are critical for best performance which is particularly important in few-shot learning. We thus propose to also meta-learn two important hyperparameters, namely learning rate and regularization weight, together with the combination weights. On the miniImageNet 1-shot learning cases, we achieved 64.3% and 71.4% recognition accuracies respectively with backbones ResNet-25 and WRN-28-10.

1.2 Incremental/Continual learning

Natural learning systems such as humans inherently work in an incremental manner as the number of concepts increases over time. They naturally learn new concepts while not forgetting previous ones. In contrast, current machine learning systems, when continuously updated using novel incoming data, suffer from catastrophic forgetting (also called catastrophic interference), as the updates can override knowledge acquired from previous data. Catastrophic forgetting, thus, becomes a major problem for continual learning systems. Our solution falls in two aspects – data storage and network architecture.

Learning to optimize exemplars. One intuitive way of continual learning is re-training with both old and new classes. However, it is neither desirable nor scalable to retain the entire data of old classes since it is too costly, and thus existing methods restrict the number of exemplars that can be kept around. E.g., only 20 exemplars per class can be stored and passed to the subsequent training phases. We found that the class boundaries learned from few random or center exemplars are weak in later training phases. We propose to tackle this problem via automatically optimizing a set of efficient but synthesized exemplars. By doing this, we are able to efficiently improve four popular incremental learning baseline methods by large margins [Liu et al. CVPR 2020]. Our learning method is empirically proved to be low-cost and generic.

Learning to combine neural networks. Class-Incremental Learning (CIL) aims to learn a classification model with the number of classes increasing phase-by-phase. An inherent problem in CIL is the stability-plasticity dilemma between the learning of old and new classes, i.e., high-plasticity models easily forget old classes but high-stability models are weak to learn new classes. We alleviate this

issue by proposing a novel network architecture called Meta-Aggregation Networks (MANets) in which we explicitly build two residual blocks at each residual level (taking ResNet as the baseline architecture): a stable block and a plastic block. We aggregate the output feature maps from these two blocks and then feed the results to the next-level blocks. We meta-learn the aggregation weights in order to dynamically optimize and balance between the two types of blocks, i.e., inherently between stability and plasticity. We conduct extensive experiments on three CIL benchmarks: CIFAR-100, ImageNet-Subset, and ImageNet, and show that many existing CIL methods can be straightforwardly incorporated on the architecture of MANets to boost their performances [Liu et al. 2020] (under review).

2. Causal reasoning is towards highly robust AI models!

What we need in causal reasoning is a principled approach to analyze and reveal the biased components hiding in the data but causing prediction failures. Given this capability, AI models can make more robust predictions with less training data or even weakly-supervised training data (i.e., no ground truth labels).

2.1 Weakly-supervised semantic segmentation

Semantic segmentation aims to classify each image pixel into its corresponding semantic class. It is an indispensable computer vision building block for scene understanding applications such as autonomous driving and medical imaging. However, the pixel-level labeling is expensive, e.g., it costs about 1.5 man-hours for one 500x500 daily-life image. Therefore, to scale up, we are interested in Weakly-Supervised Semantic Segmentation (WSSS), where the “weak” denotes a much cheaper labeling cost at the instance-level or even at the image-level [26, 63]. In particular, we focus on the latter as it is the most economic way— only a few man-seconds for tagging an image.

In a NeurIPS 2020 oral paper [Zhang et al. NeurIPS 2020], we present a causal inference framework to improve WSSS. Specifically, we aim to generate better pixel-level pseudo-masks by using only image-level labels -- the most crucial step in WSSS. We attribute the cause of the ambiguous boundaries of pseudo-masks to the confounding context, e.g., the correct image-level classification of "horse" and "person" may be not only due to the recognition of each instance, but also their co-occurrence context, making the model inspection (e.g., CAM) hard to distinguish between the boundaries. Inspired by this, we propose a structural causal model to analyze the causalities among images, contexts, and class labels. Based on it, we develop a new method: Context Adjustment (CONTA), to remove the confounding bias in image-level classification and thus provide better pseudo-masks as ground-truth for the subsequent segmentation model. On PASCAL VOC 2012 and MS-COCO, we show that CONTA boosts various popular WSSS methods to new state-of-the-arts.

2.2 Few-shot learning

In this work [Yue et al. NeurIPS 2020], we uncover an ever-overlooked deficiency in the prevailing Few-Shot Learning (FSL) methods: the pre-trained knowledge is indeed a confounder that limits the performance. This finding is rooted from our

causal assumption: a Structural Causal Model (SCM) for the causalities among the pre-trained knowledge, sample features, and labels. Thanks to it, we propose a novel FSL paradigm: Interventional Few-Shot Learning (IFSL). Specifically, we develop three effective IFSL algorithmic implementations based on the backdoor adjustment, which is essentially a causal intervention towards the SCM of many-shot learning: the upper-bound of FSL in a causal view. It is worth noting that the contribution of IFSL is orthogonal to existing fine-tuning and meta-learning based FSL methods, hence IFSL can improve all of them, achieving a new 1-/5-shot state-of-the-art on the public benchmarks --- miniImageNet, tieredImageNet, and cross-domain CUB datasets.

Selected Publications and Outputs

[Sun et al. CVPR 2019] Meta-transfer learning for few-shot learning, by **SUN, Qianru**; LIU, Yaoyao; CHUA, Tat-Seng; SCHIELE, Bernt. (2019). 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, June 16-20, (pp. 403-412) Long Beach. (Published)

[Sun et al. PAMI 2020] Meta-transfer learning through hard tasks, by **SUN, Qianru**; LIU, Yaoyao; CHEN Zhaozheng; CHUA Tat-Seng; SCHIELE Bernt. (2020). IEEE Transactions on Pattern Analysis and Machine Intelligence, 1-14. (Advance Online)

[Li et al. NeurIPS 2019] Learning to self-train for semi-supervised few-shot classification, by LI, Xinzhe; **SUN, Qianru**; LIU, Yaoyao; ZHENG, Shibao; ZHOU, Qin; CHUA, Tat-Seng; SCHIELE, Bernt. (2019). Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, December 8, (pp. 1-11) Vancouver. (Advance Online)

[Liu et al. CVPR 2020] Mnemonics training: Multi-class incremental learning without forgetting, by LIU, Yaoyao; SU, Yuting; LIU, An-An; SCHIELE, Bernt; **SUN, Qianru**. (2020). Proceedings of the 33rd Conference on Computer Vision and Pattern Recognition, CVPR '20, Virtual Conference, June 14-19, (pp. 12245-12254). (Published)

[Liu et al. ECCV 2020] An ensemble of epoch-wise empirical bayes for few-shot learning, by LIU, Yaoyao; SCHIELE, Bernt; **SUN, Qianru**. (2020). Proceedings of the 16th European Conference on Computer Vision, ECCV 2020, Virtual, August 23-28, (pp. 404-421) Glasgow, UK: Springer. (Published)

[Liu et al. 2020] Meta-Aggregating Networks for Class-Incremental Learning, by LIU, Yaoyao; SCHIELE, Bernt; **SUN, Qianru**. arXiv preprint arXiv:2010.05063, 2020 (Under Review)

[Zhang et al. NeurIPS 2020] Causal intervention for weakly-supervised semantic segmentation, by ZHANG Dong; ZHANG, Hanwang; TANG, Jinhui; HUA, Xian-Sheng; **SUN, Qianru**. (2020). Proceedings of the 34th Conference on Neural Information Processing Systems, NeurIPS 2020, Vancouver, Canada, December 6-12, (pp. 1-12) Virtual Conference: (Published)

[Yue et al. NeurIPS 2020] Interventional few-shot learning, by YUE, Zhongqi; ZHANG Hanwang; **SUN, Qianru**; HUA, Xian-Sheng. (2020). Proceedings of the 34th Conference on Neural Information Processing Systems, NeurIPS 2020, Vancouver, Canada, December 6-12, (pp. 1-23) Virtual Conference: (Published)

NOTE: More related publications can be found at qianrusun.com.