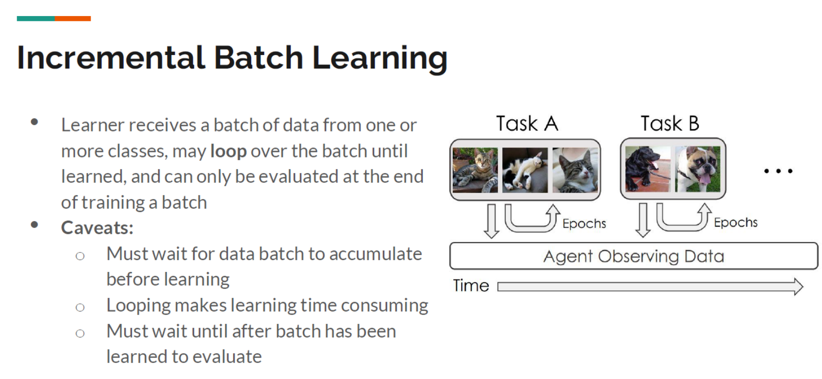
cvpr-20-conf-online

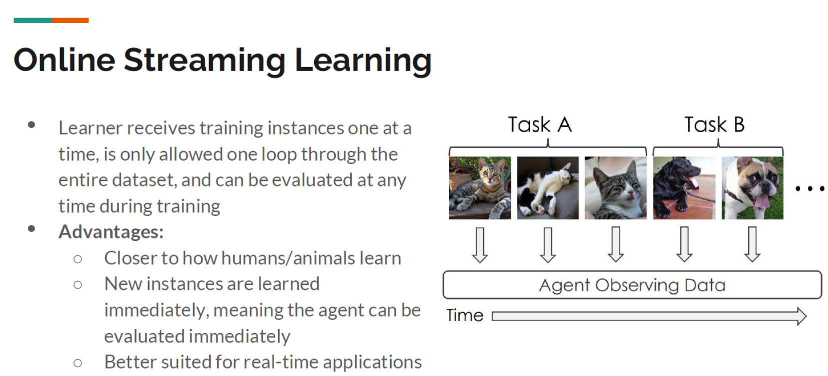
ycwang416

June 2020

1 Workshop on Continual Learning in Computer Vision

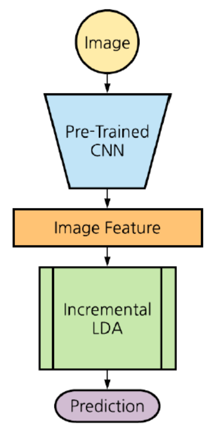
1.1 Lifelong Machine Learning with Deep Streaming Linear Discriminant Analysis (best paper)



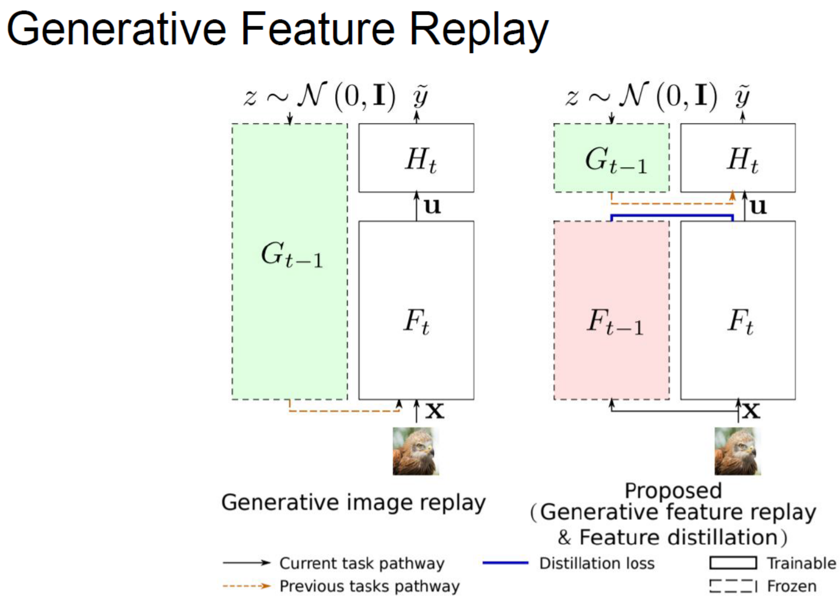


Deep Streaming Linear Discriminant Analysis

1. Extract image feature from **pre-trained deep CNN**
2. Update **class-specific running mean vector** and running **shared covariance matrix** among classes
3. During inference, a prediction is made by assigning the label of the **closest Gaussian in feature space** defined by the class mean vectors and covariance matrix



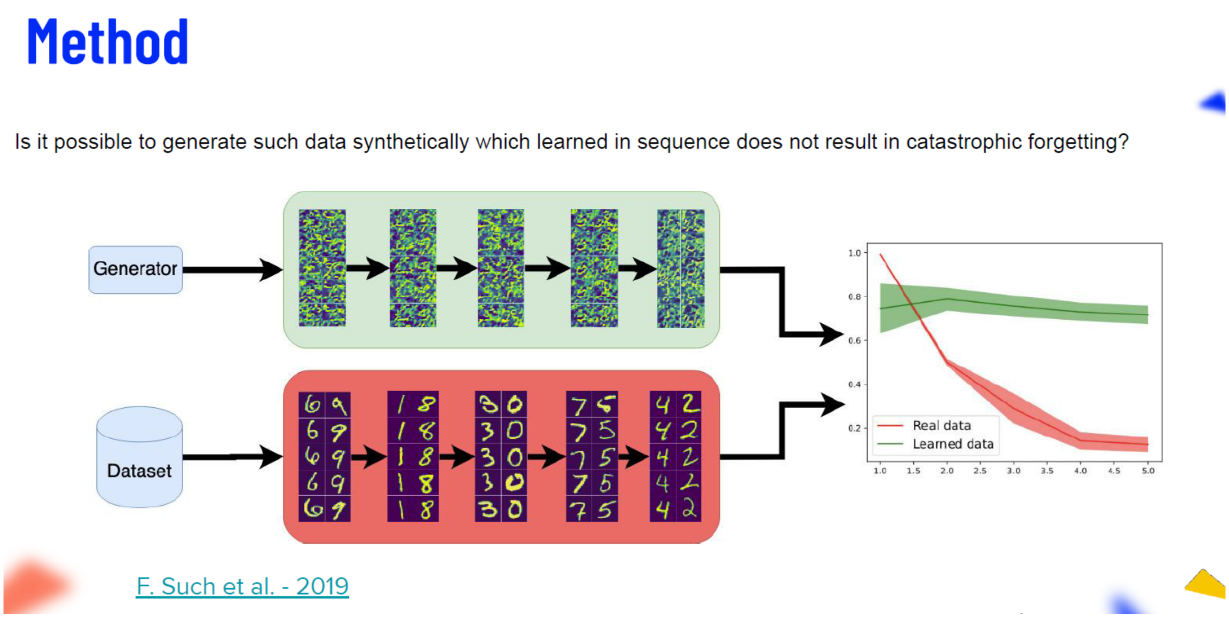
1.2 Generative Feature Replay for Class-incremental Learning



We propose generative feature replay for continual learning. Our method:

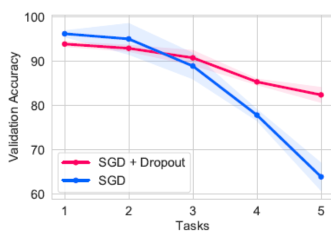
- is computationally efficient and scalable to large datasets
- outperforms other methods without exemplars by a large margin

1.3 Reducing catastrophic forgetting with learning on synthetic data

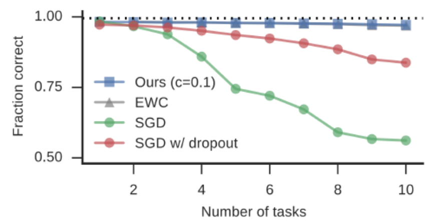


1.4 Dropout as an Implicit Gating Mechanism for Continual Learning

Observation: Networks trained with dropout tent to forget at a slower rate.

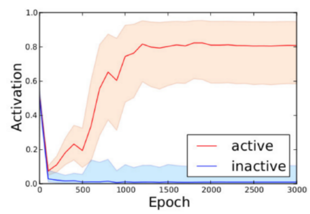


(a) Our Paper



(b) Zenke et al., 2017

Claim: Training with dropout and learning rate decay implicitly creates a gating behavior in network such that for different tasks, different paths of the network are active.



(a) The three phases of learning. For a particular input, a typical active neuron (red) starts out with low variance, experiences a large increase in variance during learning, and eventually settles to some steady constant value. In contrast, a typical inactive neuron (blue) quickly learns to stay silent.

Figure: From Baldi and Sadowski, 2013

- Dropout regularization helps to create gates in the network by pushing the neurons to be either highly active or highly inactive during the learning experience.
- When facing new tasks, the regularization mechanism will change the semi-active neurons more compared to active or inactive neurons, which helps to preserve the task-specific pathways when learning subsequent tasks.
- The learning rate decay, also helps preserving gates throughout the continual learning experience.

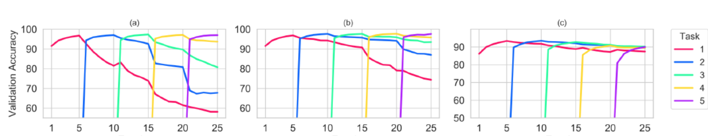


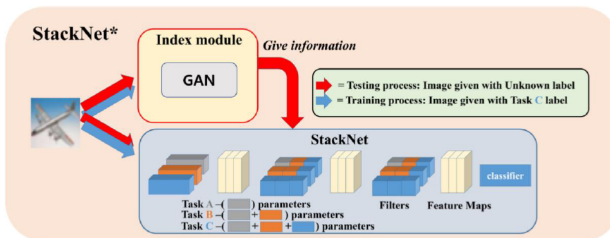
Figure: Increasing the stability and reducing the plasticity from left to right by increasing the the dropout rate and learning rate decay.

1.5 Stacking feature maps for Continual learning

Two complementary components StackNet and Index module.

Method

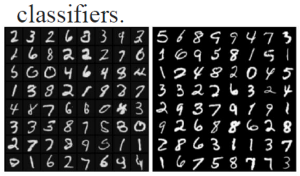
- After training the **StackNet** from the previous task with a certain amount of parameters, additional parameters in the convolutional modules are stacked and trained with the next task.
- In order to determine which combination of filters to use, we adopted a **index module** which can distinguish the origin of a given input sample.
- The combined **index module** and **StackNet** can prevent the catastrophic forgetting using different parameters for different tasks under the constraint network capacity.



StackNet* = StackNet + Index module

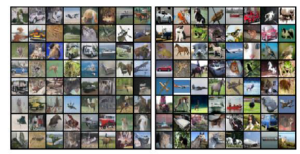
Index module (GAN)

- we introduce the index module inspired by [3] using several different methods, which is able to estimate the task of the given input data and inform this to the StackNet to specify which filters to use.
- We train a task-specific generative model using a GAN to generate pseudo-samples of each task.
- A task-wise binary classifier is trained to classify whether the given input is from the task or not.
- Index module can figure out the task of input data with maximum probability across the set of classifiers.



(a) Generated samples with the generator of index module (b) Real samples

(a) Generated samples with the generator of index module (b) Real samples

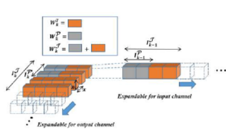


(a) Generated samples with the generator of index module (b) Real samples

[3] Li, Zhizhong, and Derek Hoiem. "Learning without forgetting." *IEEE transactions on pattern analysis and machine intelligence* 40.12 (2017): 2935-2947.

StackNet

- We propose an efficient way to allocate the capacity of a network according to the given tasks.
- While training the StackNet with several tasks, each task uses a different part of the StackNet.
- we utilize a network divided into several parts and refer it as StackNet where each part takes charge of a particular task.
- we introduce a filter index I^J (J task). The filter index I^J defines the range of filters to use for the specific task J in every layers.
- To avoid Forgetting, when we train the new weight for new task, we freeze the previous weights and only update the new weights for new task.



Algorithm 1 StackNet Training

Input: $W^{\mathcal{J}}, I^{\mathcal{P}}, I^{\mathcal{J}}$
Output: $W^{\mathcal{J}*}$
1: Freeze the $W^{\mathcal{P}} \subset W^{\mathcal{J}}$ using the information $I^{\mathcal{P}}$
2: Initialize the $W^{\mathcal{T}}$ with $W^{\mathcal{P}}$ and a random noise
3: $W^{\mathcal{J}*} \leftarrow \arg\min(\mathcal{L}_c)$
{Update $W^{\mathcal{T}}$ using backpropagation}

2 Poster

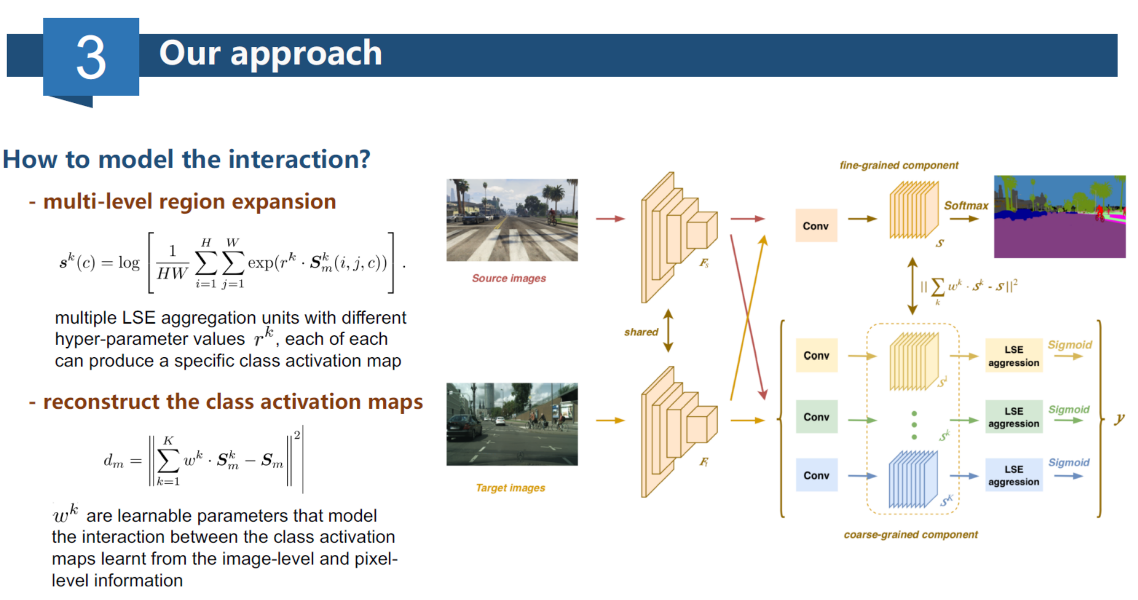
2.1 Cross-Domain Semantic Segmentation via Domain-Invariant Interactive Relation Transfer

Previous work drawbacks

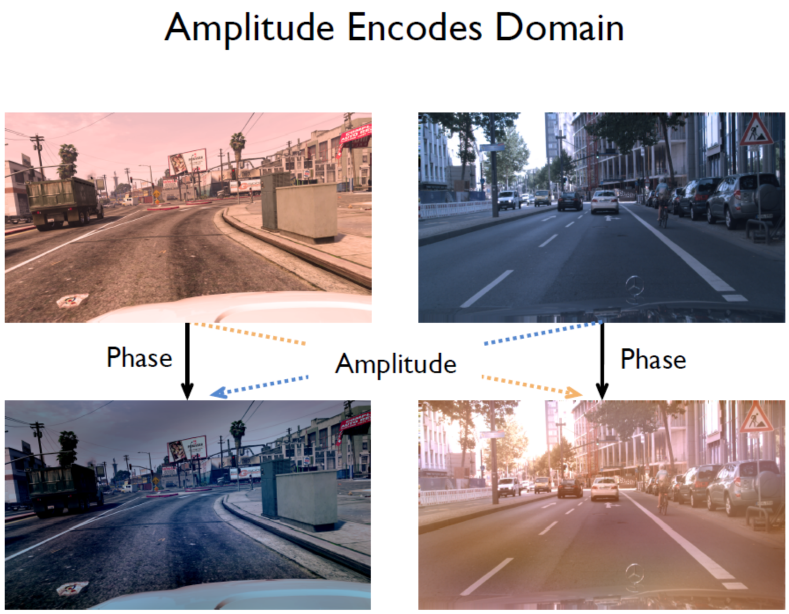
- it remains unclear what comprises the data instances in segmentation
- the minimax game is hard to solve

pivot interaction tranfer

- easy to infer the **img-level categories** of target imgs
- there exists a **strong interaction** between the img-level categories and pixel-level semantic segmentation
- this interaction is invariant across different domains in urban traffic scenes
- main idea is model the **domain-invariant relations** as pivot knowledge for transfer



2.2 Fourier Domain Adaptation for Semantic Segmentation



KEY IDEA The key idea is to do style transfer while preserving the semantic information of images. Different from GAN, the FFT method does not include any training parameters. It's very simple and efficient and achieves very good results.

FFT maps a source image to a target 'style' without changing semantic content. That is the most fascinating and crucial part for domain adaption because pervious cycle-GAN based methods will alter the semantic information more or less.

Main Steps: Do FFT for source and target images. Swap the low frequency part of the amplitude of source and target images. Map back images to original image space without changing the phase part.

The semantic details of source images are well preserved but resemble the appearance of target domain. Based on this, the model can make use of the label information from source to train the decision boundary of target domain.

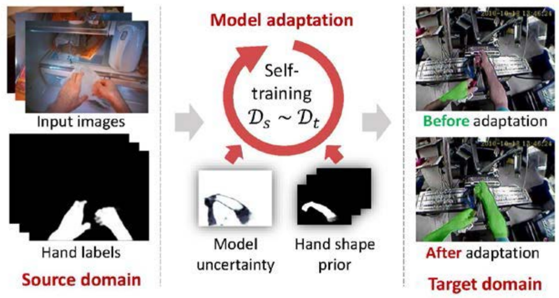
The second part innovation is that it generates more confident pseudo-labels by training multiple models with different FFT band parameters.

Experiment of GTA5, Cityscapes, Synthia show competitive results but this method is very simple and does not require any learning.

2.3 Generalizing Hand Segmentation in Egocentric Videos With Uncertainty-Guided Model Adaptation

Keywords: Domain adaptation, Uncertainty estimation, Real-time hand segmentation.

Problem:



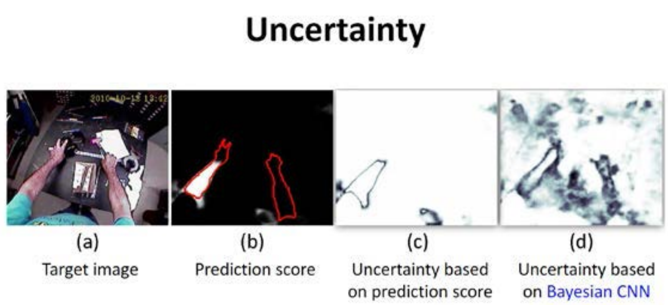
Goal

- Adapt a hand segmentation model pre-trained on a source domain to a new target domain without labels.
- Ideal uncertainty should reflect the models' "real" confidence about its prediction

Key Idea

- Use uncertainty to guide self-training with pseudo-labels in the target domain.
- Use pre-trained hand discriminator to enforce hand shape consistency.
- The model adapts quickly to the target domain with a few unlabeled samples.

Results

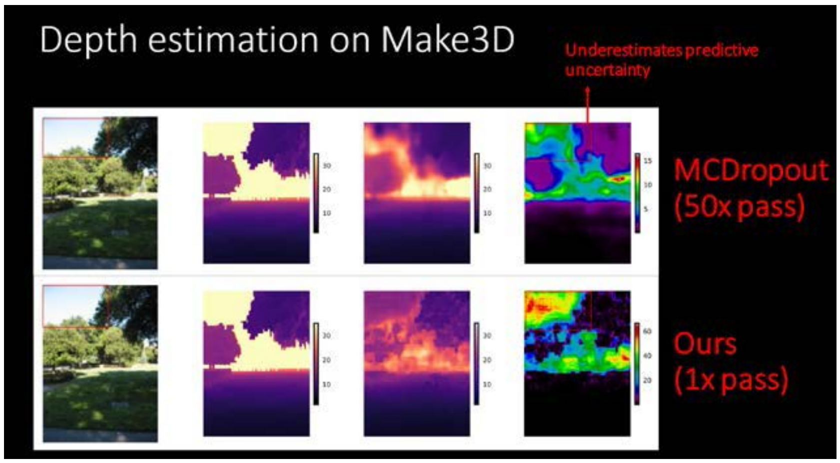


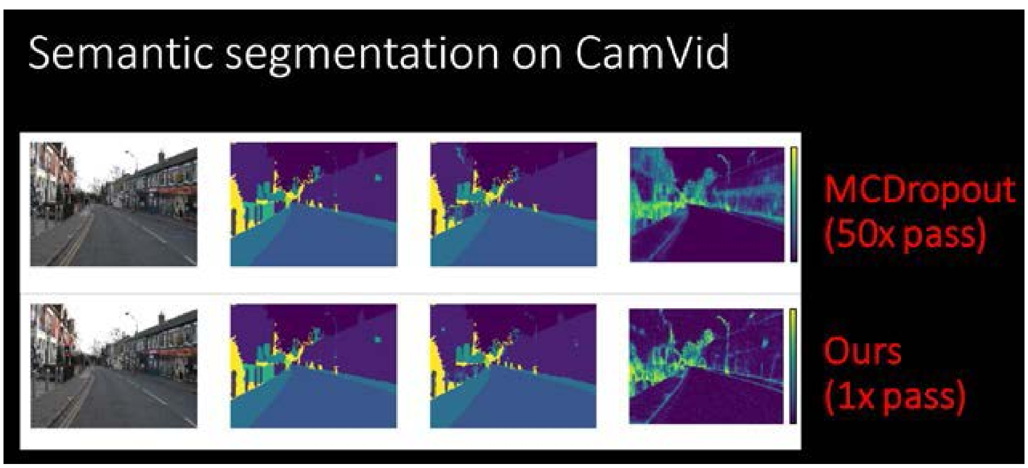
2.4 Scalable Uncertainty for Computer Vision with Functional Variational Inference

Keywords: Uncertainty estimation,

Goal:

- Monte Carlo Dropout (MC-Dropout) has been popular in Computer Vision, but requires multiple forward passes at test time.
- Our model does Bayesian DL with only one forward pass.



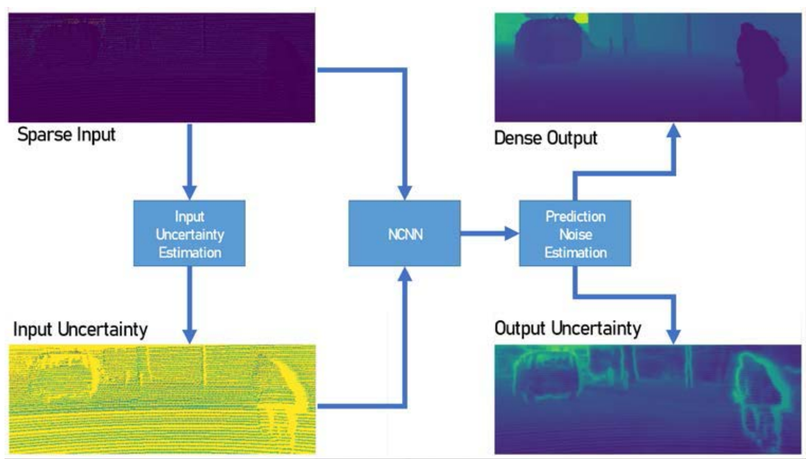


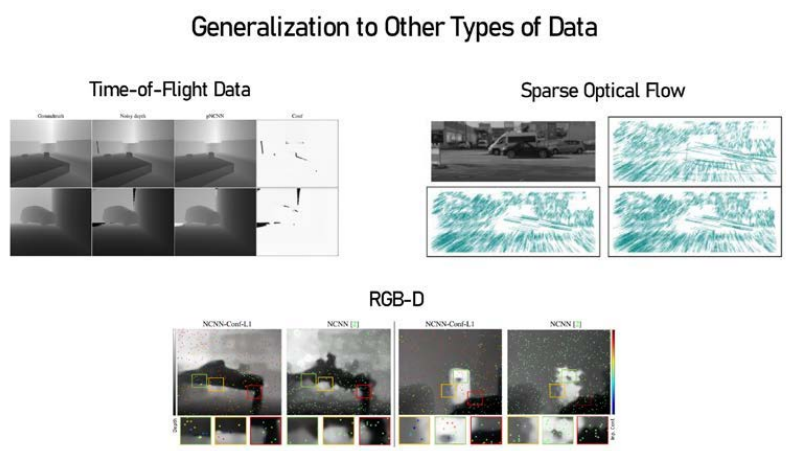
2.5 Uncertainty-Aware CNNs for Depth Completion: Uncertainty from Beginning to End

Keywords: Depth data, Uncertainty estimation **Goal:**

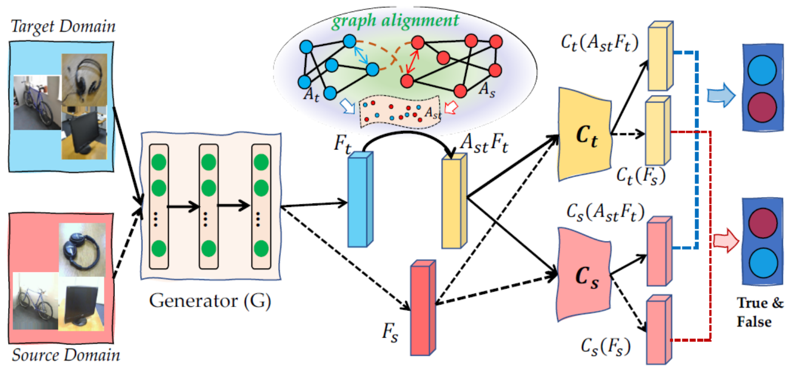
- Depth data is challenging due to sparsity and uncertainty.
- Our method achieves a real-time, fully-unguided (no RGB images needed), sparsity-aware and uncertainty-aware probabilistic Normalized Convolutional Neural Networks (pNCNN).

Key idea:





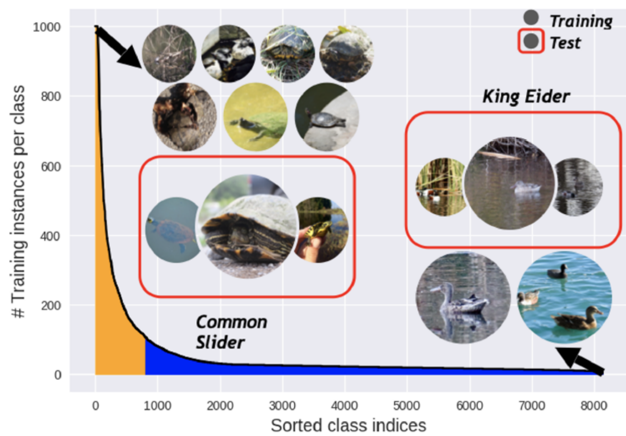
2.6 Structure Preserving Generative Cross-Domain Learning



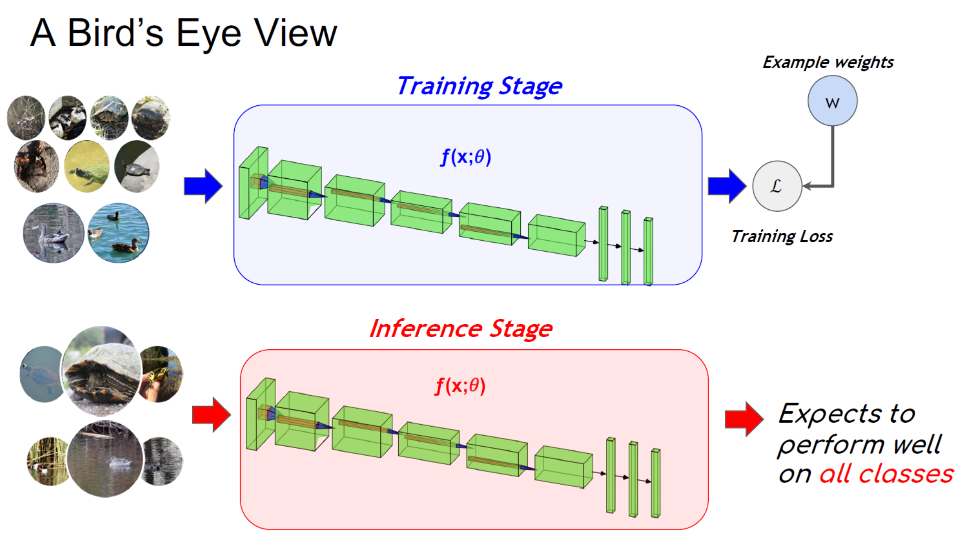
Features F_s and F_t are extracted from raw data through generator (ResNet). Matching relationship of two domains is captured according to graph distribution, and two classifiers are built in adversarial learning.

Source-supervised classifier are fed with the combine of matching relation and features from target domain. A symmetric adversarial manner is explored to train two domain-specific classifiers.

2.7 Rethinking Class-Balanced Methods for Long-tailed Visual Recognition from a Domain Adaptation Perspective



New Perspective - Domain Adaptation
Setup:
Source domain(labeled data): $D_s = \{x_m, y_m\}_{m=1}^{M}$ $P_s(X, Y)$
Target domain (no labels for training): $D_T = \{(x_n, ?)\}_{n=1}^{N}$ $P_T(X, Y)$
Existing work Assume the target shift
$P_s(x|Commonslider) = P_t(x|Commonslider)$ $P_s(x|KingEider) = P_t(x|KingEider)$
But $P_s(x|Commonslider) = P_t(x|Commonslider)$ $P_s(x|KingEider) \neq P_t(x|KingEider)$



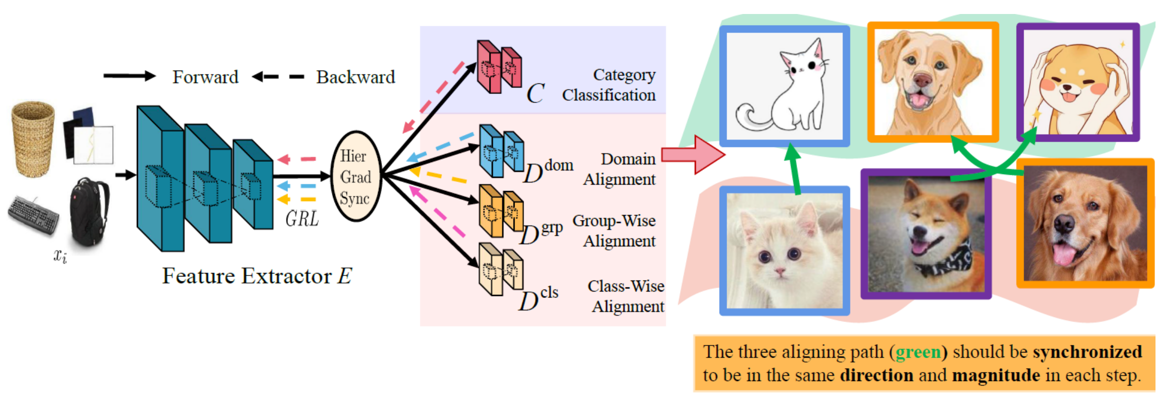
2.8 Unsupervised Domain Adaptation with Hierarchical Gradient Synchronization

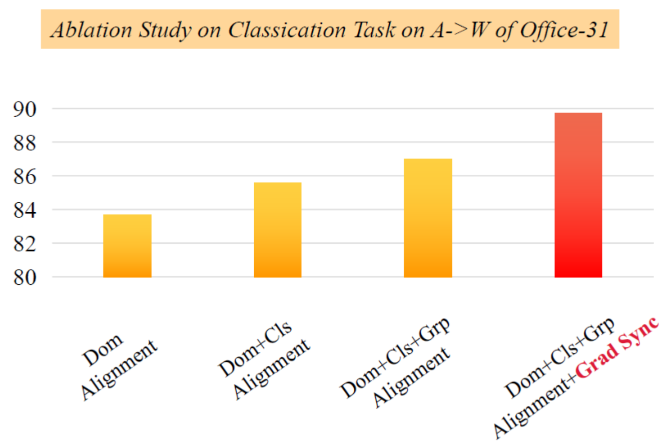
Limitation of existing methods

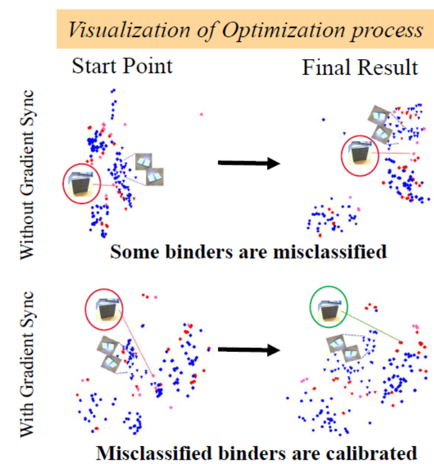
- Independently aligning **global** and **local** distributions
- Global alignment: Ignore fine category information-¿exist misaligned categories
- Local alignment: Ignore global information-¿overfit to the same category
- **Compromise** of simply combining them together

METHOD

- Make the global and local alignments consistent
- **Gradient Synchronization:** make the global and local alignment gradients synchronization on direction and magnitude
- Obtain better performance on commonly used UDA datasets







Discovering the **relation** between global and local alignments, e.g., **gradient synchronization**, shows its rationality for UDA problem and deserves more exploration.