Abstract

Deep (machine) learning in recent years has significantly improved the predictive modeling strengths for applications in the areas of computer vision, speech recognition, and natural language processing. In deep learning, weight sharing is the driving force behind the successes of many deep neural networks. In weight sharing, the neural network weights are shared among different local parts/pieces of the data or of a network component. These techniques in general can boost learning efficiency and effectiveness, improving regularization for enhanced generalization performance. In this thesis, we introduce weight sharing techniques to tackle a range of problems in visual recognition tasks. Specifically, our contributions are novel methods to share deep neural network weights among (i) different image sub-regions, (ii) pre-trained classification weights of different object classes, (iii) different succeeding network layers, and (iv) differently downsampled convolutional network feature maps. Throughout the thesis, we demonstrate how each of the proposed methods can address a specific visual recognition problem.

Firstly, we apply weight sharing to improve the saliency detection performance of deep networks. A coarse saliency map of the input image's salient object is initially predicted by a convolutional-deconvolutional network (or equivalently a fully convolutional network). Following that, the coarse saliency map is sequentially refined by a separate network called Recurrent Attentional Convolutional-Deconvolutional Network (RACDNN) that shares its network weights across its sequential steps. At every sequential (temporal) step, the same RACDNN is applied to a newly selected/attended image sub-region and it performs saliency map refinement for that particular sub-region. Weight sharing is applicable because different image subregions are considered as different instances of the same data. Despite having a large number of learnable network parameters and no access to additional training data, RADCNN can be trained well and it outperforms the "single-step"

(coarse saliency detection) baseline and state-of-the-art saliency detection methods by large margins, on several saliency detection benchmark datasets.

Secondly, we introduce class-to-class weight transfer networks (WTN). Class-to-class WTN makes use of weight sharing to effectively transfer object class knowledge from the classification weights of a pretrained large-scale image classification network, to the classification weights of an object detection network. The knowledge transfer is to expand the number of classes handled by the object detection network, given only a detection training dataset with bounding box annotations covering a smaller number of classes. The network weights of class-to-class WTN are shared across the different object classes, such that class-to-class WTN learns a generic class-agnostic mapping function to adapt the per-class weights from classification task to detection task. The learned generic mapping can then be applied to any novel object classes which have not been seen by the detection network during training. A loss function is proposed to encourage class-to-class WTN to maintain the linear combination relationships between novel and common classes for improved detection performance on novel classes. Experiments on large-scale detection datasets validate the effectiveness of class-to-class WTN on both seen and unseen classes.

Thirdly, we consider *soft* weight sharing among different composite network layers to facilitate cross-layer information inflows for enhanced representational capability of deep convolutional networks in the context of image classification. This gives rise to a new convolutional network architecture called Deluge Networks (DelugeNets) which pass the output (feature activations) of any preceding composite network layer as direct inputs to all succeeding composite layers. For the incoming feature activations of preceding composite layers, the succeeding composite layers first apply rather lightweight cross-layer depthwise convolutional weights on them for feature aggregation. The heavy lifting is carried out by the last convolutional weight layer in each preceding composite layer that performs cross-channel

convolutions. It can be seen as decomposing the convolutional weights into two parts: a heavy one found in the preceding layer and shared by all succeeding layers, and another lightweight one associated with the different succeeding layers. The proposed soft weight sharing technique allows very deep DelugeNets to perform information transfer across a huge number of composite layers in an effective and efficient manner. Experiments on small-scale and large-scale datasets show that DelugeNets perform competitively to state-of-the-art deep convolutional networks at lower computational and parameter costs.

Fourthly, inspired by weight sharing, we propose Stochastic Downsampling (SDPoint), a method to overcome a major limitation suffered by conventional convolutional networks – their computational costs during inference are fixed and cannot be changed. SDPoint downsamples the feature maps at a randomly selected point (layer index) in the network hierarchy during training. The many downsampling configurations known as SDPoint instances each entail a unique downsampling point and downsampling ratio. The same convolutional network weights are shared across the different SDPoint instances that downsample feature maps differently. During inference, the computational cost of a network trained with SDPoint can be instantaneously adjusted to fit a given computational budget, by manually selecting an appropriate SDPoint instance. Additionally, sharing weights between different feature map scales provides significant regularization benefits, making the networks trained with SDPoint less sensitive to scales and more competent. In image classification experiments, SDPoint demonstrates significant cost-accuracy and computational advantages, over independently-trained (non-weight sharing) baselines with various inference costs.

Author Publications

- **Jason Kuen**, Xiangfei Kong, Zhe Lin, Gang Wang, Jianxiong Yin, Simon See, Yap-Peng Tan. Stochastic Downsampling for Cost-Adjustable Inference and Improved Regularization in Convolutional Networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

- **Jason Kuen**, Xiangfei Kong, Gang Wang, Yap-Peng Tan. DelugeNets: Deep Networks with Efficient and Flexible Cross-layer Information Inflows. *International Conference on Computer Vision Workshop (ICCVW) on Compact and Efficient Feature Representation and Learning in Computer Vision (CEFRL),* ***oral paper*** (2017).

- **Jason Kuen**, Zhenhua Wang, Gang Wang. Recurrent Attentional Networks for Saliency Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).

- Ping Hu, Gang Wang, Xiangfei Kong, **Jason Kuen**, Yap-Peng Tan. Motion-Guided Cascaded Refinement Network for Video Object Segmentation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

- Jianlou Si, Honggang Zhang, Chun-Guang Li, **Jason Kuen**, Xiangfei Kong, Alex Kot, Gang Wang. Dual Attention Matching Network for Context-Aware Feature Sequence based Person Re-Identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

- Jiuxiang Gu*, Zhenhua Wang*, **Jason Kuen**, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, Tsuhan Chen. Recent Advances in Convolutional Neural Networks. *Pattern Recognition (PR)* (2017).