Deep Convolutional Sum-Product Networks

Cory J. Butz
butz@cs.uregina.ca
University of Regina
Canada

Jhonatan S. Oliveira
oliveira@cs.uregina.ca
University of Regina
Canada

André E. dos Santos
dossantos@cs.uregina.ca
University of Regina
Canada

André L. Teixeira
teixeira@cs.uregina.ca
University of Regina
Canada

Abstract

We give conditions under which *convolutional neural networks* (CNNs) define valid *sum-product networks* (SPNs). One subclass, called *convolutional* SPNs (CSPNs), can be implemented using tensors, but also can suffer from being too shallow. Fortunately, tensors can be augmented while maintaining valid SPNs. This yields a larger subclass of CNNs, which we call *deep convolutional* SPNs (DCSPNs), where the convolutional and sum-pooling layers form rich directed acyclic graph structures. One salient feature of DCSPNs is that they are a rigorous probabilistic model. As such, they can exploit multiple kinds of probabilistic reasoning, including *marginal* inference and *most probable explanation* (MPE) inference. This allows an alternative method for learning DCSPNs using vectorized differentiable MPE, which plays a similar role to the generator in *generative adversarial networks* (GANs). Image sampling is yet another application demonstrating the robustness of DCSPNs. Our preliminary results on image sampling are encouraging, since the DCSPN sampled images exhibit variability. Experiments on image completion show that DCSPNs significantly outperform competing methods by achieving several state-of-the-art *mean squared error* (MSE) scores in both left-completion and bottom-completion in benchmark datasets.

Introduction

Generative models are of current interest in the deep learning community, including *generative adversarial networks* (GANs) (Goodfellow et al. 2014), variational autoencoders (Kingma and Welling 2014), neural autoregressive distribution estimators (Larochelle and Murray 2011), pixel recurrent neural networks (Oord, Kalchbrenner, and Kavukcuoglu 2016), and convolutional arithmetic circuits (Sharir et al. 2018). *Convolutional neural networks* (CNNs) (Goodfellow, Bengio, and Courville 2016) can be used in GANs. *Sum-product networks* (SPNs) (Poon and Domingos 2011) are a generative model that have received limited attention from the deep learning community (Peharz et al. 2018). An SPN is a *directed acyclic graph* (DAG), where leaf nodes are tractable distributions and each internal node is either a sum or product operation. A *valid* SPN defines a joint probability distribution and allows for efficient inference (Poon and Domingos 2011).

Conditions are given as to when subclasses of CNNs define valid SPNs, including convolutional layer filters of certain sizes and non-overlapping windows in sum-pooling layers. Satisfaction of these conditions yields a subclass of CNNs, called *convolutional* SPNs (CSPNs). CSPNs permit a vectorized representation allowing for exploitation of tensor libraries such as Tensorflow, but they also can suffer from being too shallow, and it is known that deep SPNs are more expressive than shallow SPNs (Delalleau and Bengio 2011).

We introduce *deep convolutional sum-product networks* (DCSPNs). DCSPNs permit the convolutional and sum-pooling layers to form rich DAG structures by augmenting layer tensors under conditions that maintain *decomposability* and *completeness*. As a decomposable and complete SPN is a valid SPN, our main result is that DCSPNs are a larger subclass of CNNs that define valid SPNs. DCSPNs are a rigorous probabilistic model. As such, they can exploit probabilistic reasoning, including *marginal* inference and *most probable explanation* (MPE) inference. This allows an alternative method for learning DCSPNs using vectorized differentiable MPE. We show how to vectorize MPE using a mask algorithm and how it plays a role similar to the GANs generator. Image sampling is yet another application demonstrating the robustness of DCSPNs. This involves a minor modification to the mask algorithm. Our preliminary results on image sampling are promising, since the DCSPN sampled images exhibit variability. Experimental results on left- and bottom-completion like those in Table 1 show DCSPNs achieve state-of-the-art by building deeper structures using both vertical and horizontal sum-pooling windows, which leverage local structure in the image data in both directions. Applying a simple low pass filter as a post-processing smoothing operation lowers the *mean squared error* (MSE) score from **455** to **401** for left-completion in Olivetti.

Table 1: Mean squared error (MSE) scores in Olivetti Face.

	left	bottom
P&D (Poon and Domingos 2011)	942	918
ICNN (Amos, Xu, and Kolter 2017)	833	-
D&V (Dennis and Ventura 2012)	779	782
DCGAN (Yeh et al. 2017)	935	707
DCSPN	**455**	**503**