Network in Network

Link

arxiv

Summary

- The convolution filter used in CNN is a generalized linear model for the underlying data patch. It might be sufficient when the latent concepts are linearly separable but the concepts are often not linearly separable. In CNN this is compensated by using many filters on the same layer where each filter is assumed to be accountable for capturing different variation in the concept. However having too many concepts poses a problem for the next layer which has to take into account all possible linear combination of those features. So they replace the convolution filters with a micro network which is a MLP. Just like in convolution the MLP weights are shared and feature is extracted by sliding the network over the image. Since MLP are universal function approximator they should be able to capture latent concepts better than plain convolution.
- In CNN, convolution feature extractors are followed by fully connected layers which act as the classifier. However fully connected layers are prone to overfitting and hard to interpret. In NIN the fully connected layer is replaced by a global average pooling layer and the pooled vector is fed into softmax. It preserves correspondence between feature maps and categories. There is no risk of overfitting since there is no parameter to optimize. It is also more robust to spatial translation since the features are summed. Experiment shows that global average pooling can act as a regularizer. When no dropout is used, the fully connected layer performs poorly compared to the global average pooling in NIN.
- Maxout network performs maximum pooling over the feature maps and so it can generalize among different concepts. However it requires the concepts to lie within a convex set in the input space. NIN doesn't require such constraint.