

Advanced Training Techniques for Convolutional Neural Networks

Mohammed Luqmaan Akhtar

January 7, 2026

Data Augmentation

Data augmentation increases dataset diversity and reduces overfitting.

- **Rotation:** Random rotations to change viewpoint.
- **Flipping:** Horizontal and vertical flips help in expanding data set.
- **Cropping:** Random cropping forces models to focus on local features.
- **Color Jittering:** Random changes in brightness and contrast improve generalisation of model.

Transfer Learning

Transfer learning is a machine learning technique where a model trained on one task is repurposed as the foundation for a second task. This approach is beneficial when the second task is related to the first or when data for the second task is limited.

Using learned features from the initial task, the model can adapt more efficiently to the new task, accelerating learning and improving performance. Transfer learning also reduces the risk of overfitting, as the model already incorporates generalizable features useful for the second task.

The process starts by selecting a pre-trained model. This is known as the base model. The lower layers in any model capture general features and the higher layers identify the complex ones. We decide which layers to freeze and which layers to modify for our specific task and fine-tune the layers that needs to be modified.

There are two types of fine-tuning strategies:

- **Partial Fine-Tuning:** Only higher layers are trained while lower layers remain frozen.
- **Full Fine-Tuning:** All layers are updated during training.

Batch Normalization in CNNs

Batch Normalization normalizes the activations of a layer across a mini-batch to have zero mean and unit variance. In CNNs, it is typically applied after the convolution layer and before the activation function.

This reduces internal covariate shift, allowing the network to train faster and use higher learning rates. Batch normalization also provides a regularization effect, reducing overfitting. It improves training stability in deep CNN architectures.

Dropout in CNNs

Dropout is a regularization technique that randomly disables a fraction of neurons during training. In CNNs, dropout is commonly applied after fully connected layers rather than convolution layers. This prevents the network from relying too heavily on specific neurons, reducing overfitting and improving generalization.

Degradation Problem

The degradation problem refers to the phenomenon where adding more layers to a deep neural network leads to higher training error instead of improved performance. This occurs not due to overfitting, but because very deep networks become difficult to optimize. Gradients struggle to propagate through many layers, making learning ineffective. Residual networks (ResNets) address this problem using skip connections that allow direct gradient flow.