Regularization and Transfer Learning

Discussion Questions

1. What are the different techniques that can be used to address overfitting in CNN models?
2. What is Transfer Learning?
3. When can Transfer Learning be used?
4. Is using Transfer Learning better than building a CNN model from scratch?

Techniques to address overfitting - Data Augmentation

✓ What is Data Augmentation ?

- Data augmentation is a set of techniques for artificially increasing data volume by generating new data points from existing data.

- Some of the simple transformations applied to the image are - **Geometric transformations** such as Flipping, Rotation, Translation, Cropping, Scaling, **and Color space transformations** such as color casting, varying brightness, and noise injection, **Random Erasing** e.t.c.

Benefits of Data Augmentation

1. Increasing the amount of data
2. Adding diversity to the training data
3. Addressing the class imbalance problem in classification tasks

Techniques to address overfitting - Data Augmentation

Original Image

Geometric Transformation

Colour space Transformation

Random Erasing

Techniques to address overfitting - Batch Normalization

✓ What is Batch Normalization ?

- Batch normalisation is a technique for standardizing network inputs that can be applied to either the activations of a previous layer or to the inputs directly.

- It stabilizes the learning process and significantly reduces the number of training epochs by standardizing the layer's inputs for each mini-batch.

Tips for using Batch-Norm

1. Batch normalisation can be applied to the layer's inputs **before** activation function in the previous layer (e.g, ReLU)

2. It may require the usage of considerably **higher learning rates**, which may further accelerate the learning process to stabilize the network.

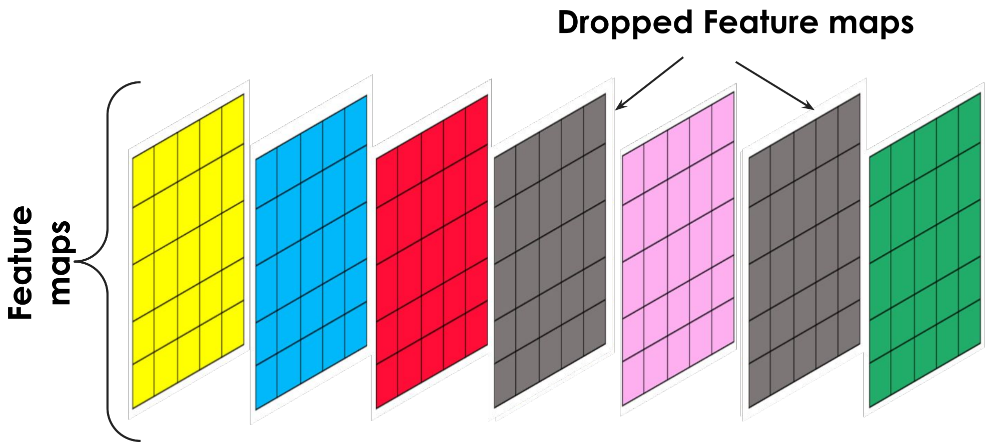
3. Batch normalisation has a regularisation effect, which reduces generalisation error and may **eliminate the need for dropout for regularisation.**

Techniques to address overfitting - Spatial Dropout

✓ What is Spatial Dropout ?

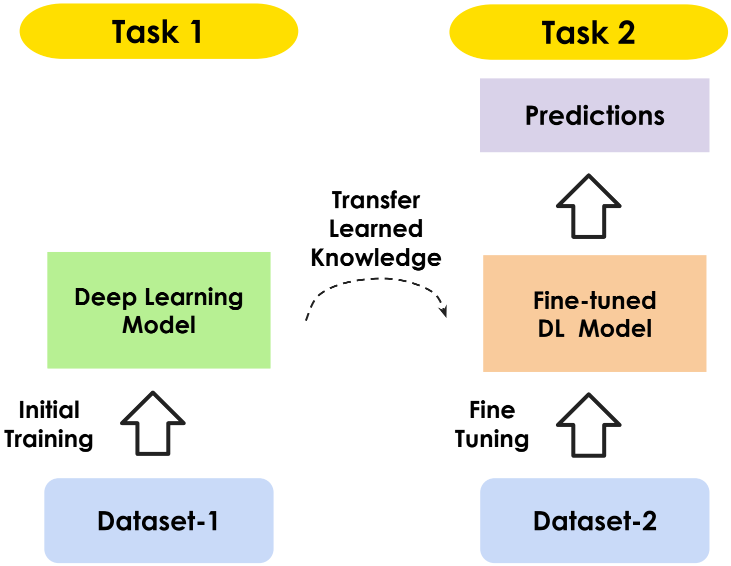
- An alternative way to use dropout with convolutional neural networks is to drop entire feature maps from the convolutional layer. This is called **Spatial Dropout**.

- It prevents the co-adaptation of pixels with their neighbors across feature maps and forces them to learn as if no other feature maps exist. Thus, it promotes independence between feature maps.



What is Transfer Learning ?

- If you are using a specific CNN architecture that has been trained before, you can use these **pretrained parameters/weights instead of random initialization** to solve your problem. It can help you boost the performance of the CNN.
- The pretrained models might have trained on large datasets like ImageNet, and taken a lot of time to **learn those parameters/weights with optimized hyperparameters**. This can save you a lot of time.
- If you have enough data, you can **fine-tune all of the layers in your pretrained network**, but don't randomize the parameters; instead let the network learn from them.



When can Transfer Learning be used ?

- Transfer Learning is beneficial when -
  - we have limited amount of labeled training data
  - we have limited amount of computational power
- Transfer learning involves training an initial network on a base dataset and task and then transferring the acquired features to a second target network for training on a target dataset and task. This method is more likely to succeed if the features are general, that is, applicable for both the base and target tasks.

How to use Transfer Learning?

1. Select a pre-trained source model from available models.
2. The pre-trained model can then be used to build a model for the second task of interest.
3. The model has to be tuned based on the input-output data available for the task

greatlearning
Power Ahead

Happy Learning !