CS231N: Convolutional Neural Networks for Visual Recognition

Cervix Type Classification (Kaggle)

Sebastiano Bea, Kevin Poulet, Dan Zylberglejd

Stanford University

Background

- Cervical cancer is easy to prevent if detected in early stages
- Different treatments depend on different physiological types
- Treatments for the right type of cervix are highly effective
- Lack of skilled labor to identify cervix type
- Advanced computer vision techniques can enhance this task
- Challenge provided by partnership between Kaggle and MobileODT, and hosted at Kaggle

Problem Statement

- Main goal: classify the type of cervix based on a single image
- There are 3 cervix types (types 1, 2 and 3)
- Evaluation is based on prediction accuracy
  - accuracy = (correct predictions) / (total cases)
- For training, use cross-entropy loss
- Use deep learning techniques (namely CNNs)

Datasets

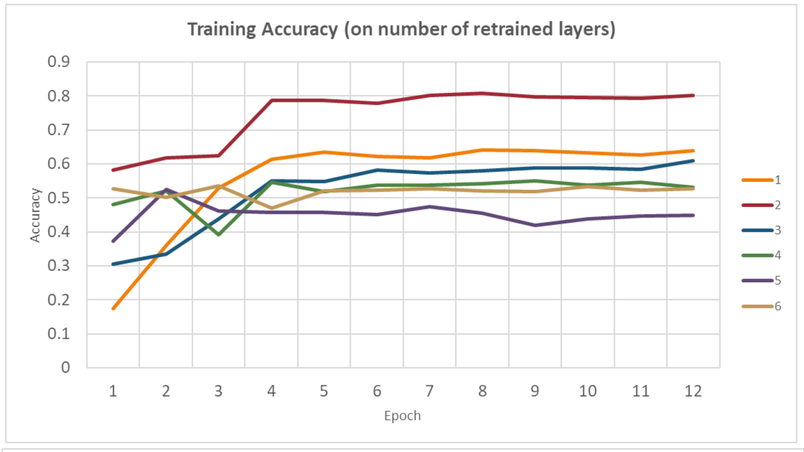
- 1481 image
  - 250 from type 1; 781 from type 2; 450 from type 3
- Split into train (60%), validation (20%) and test (20%) sets
- Sample images:

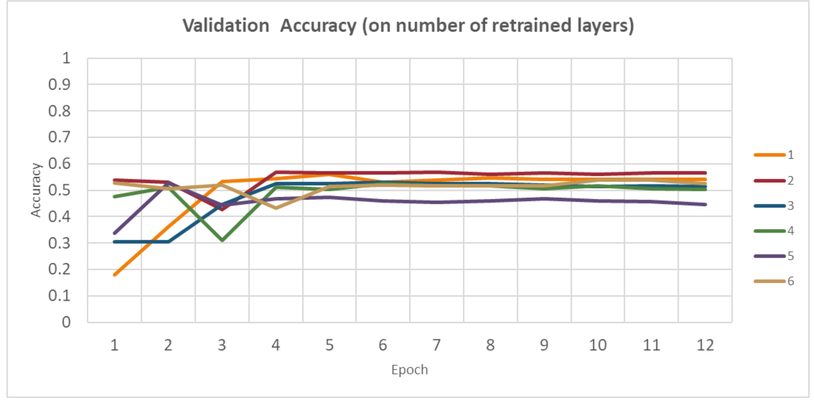
Type 1 | Type 2 | Type 3

Methods

- We base our main architecture on a pre-trained VGG16
- Fix the weights for the architecture trained on the COCO dataset for some layers
- Fine tune, updating from scratch, some of the layers
- Use validation errors to decide which layers should be fixed and which should be fine tuned

Experimental Evaluation





Current Results

- Retraining 2 layers provides the best result in training and validation accuracy
- Had to finetune not to get in local minima
- Most of the training is set at epoch 5
- Results on test set: 58%
- Test set result in line with validation set accuracy

Future Directions

- Possibility to expand the training set with additional data
- VGG is not state-of-the-art anymore
  - Use more sophisticated networks such as Resnet, Inception
- Use more sophisticated optimization methods, such as Adagrad or Adam
- Use more sophisticated initializations, such as Xavier
- Use an ensemble of models
- Understand the limits of the computer vision task; i.e., even for a human the task seems difficult, so evaluate the limits of how well an automated algorithm can perform.